

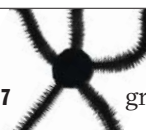
THIS WEEK

EDITORIALS

WHALES The lunge feeding of greatest ocean predators explained **p.416**

WORLD VIEW Rio meeting must not rush to set planetary boundaries **p.417**

BRITTLE STARS Bilateral steps of the five-legged groove machine **p.419**



A charter for geoengineering

A controversial field trial of technology to mitigate climate change has been cancelled, but research continues. A robust governance framework is sorely needed to prevent further setbacks.

Geoengineering research has a problem. That much should be clear following last week's cancellation of a field trial for the Stratospheric Particle Injection for Climate Engineering (SPICE) project. The solutions to this problem are not so obvious, but they must be found — and fast.

The SPICE field trial was supposed to involve spraying water into the atmosphere at an altitude of 1 kilometre using a balloon and hose-pipe, as part of a host of work exploring whether it is possible to mitigate global warming by introducing particles into the stratosphere to reflect some of the Sun's energy away from Earth.

But the field trial — which is only a small part of the overall SPICE project — became bogged down in protests and delays almost as soon as it was announced. Last week, as first reported by *Nature*, the project's lead investigator announced that it was being abandoned, citing concerns about intellectual-property rights, public engagement and the overall governance regime for such work.

Colleagues have leapt to the defence of the SPICE team, and praised its decision to continue with the theoretical strands of its work. Indeed, the researchers have acted with commendable honesty. But the SPICE issue is a perfect example of the problems that will persist until geoengineers grasp the nettle of regulation and oversight.

We have been here before. Work on 'fertilizing' the oceans to promote blooms of phytoplankton that would lock up carbon dioxide ran into similar protests and governance wrangles. In 2009, an experiment to test the idea by dumping tonnes of iron sulphate into the Southern Ocean caused huge public disquiet and went ahead only after further discussions.

Researchers argue that 'geoengineering' is a falsely inclusive term. They say that SPICE-style 'solar-radiation management' is completely different from ocean fertilization, and different again from carbon capture. But these technologies have similar aims and, when it comes to rules and regulations, they probably need to be dealt with together.

The geoengineering community has tried to bring some discipline to the emerging field. The 'Oxford Principles' — developed in 2010 by researchers at the University of Oxford, UK — offer some useful ground rules. They say that geoengineering should be regulated as a public good; there should be public participation in decision-making; research should be disclosed and results published openly; impacts should be assessed independently; and decisions to deploy the technologies should be made within a robust governance framework.

These are excellent principles. But they are vague, and cannot serve as a guide to conducting specific experiments in such a broad field.

A meeting of geoengineers in Asilomar, California, in 2009 — influenced by a meeting at the same location in 1975, when researchers hashed out guidelines for genetic engineering — produced similarly vague recommendations, such as the need to conduct research openly and to consult the public when planning research. It also called for

governments to "when necessary, create new mechanisms for the governance and oversight of large-scale climate engineering research activities".

The SPICE fiasco starkly demonstrates the need for such mechanisms. For a project of such high profile to founder on problems of

"Problems will persist until geoengineers grasp the nettle of regulation and oversight."

intellectual property, regulation or public protest would be bad enough. That it ran into difficulties in all three areas shows an underlying problem.

Of the issues raised, intellectual property may turn out to be the easiest to resolve (see page 429). Science has a long and generally happy relationship with patents, including those for technology with the ability to drive worldwide change. Likewise, lessons on public engagement and dealing with protests can be taken from earlier rows over genetic modification, stem cells, fertility work and animal research.

More troubling is the lack of an overarching governance framework. Although the SPICE trial has been cancelled, other tests of geoengineering technology will surely follow. Other work, such as fiddling with clouds to make them more reflective or to try to bring on rain, touches on both climate-change mitigation and weather modification.

Geoengineers should keep trying. They should come together and draft detailed, practical actions that need to be taken to advance governance in the field. Regulation in these cutting-edge and controversial areas needs to be working before the experiments begin, rather than racing to catch up. ■

In from the cold

Research into the Fukushima meltdowns has given a new lease of life to a valuable group.

The United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) is a relic of the cold war. Established in 1955 to study fallout from above-ground nuclear-weapons tests, the committee, which is based in Vienna, acted as one of the few channels for cooperation between the United States and the Soviet Union, and served to exchange information between East and West.

It was invaluable in its time: after the catastrophic meltdown at the Chernobyl nuclear power plant in Ukraine in 1986, the committee's ties to the Soviet Union allowed it to produce some of the first

THIS WEEK

EDITORIALS

WHALES The lunge feeding of greatest ocean predators explained **p.416**

WORLD VIEW Rio meeting must not rush to set planetary boundaries **p.417**

BRITTLE STARS Bilateral steps of the five-legged groove machine **p.419**



A charter for geoengineering

A controversial field trial of technology to mitigate climate change has been cancelled, but research continues. A robust governance framework is sorely needed to prevent further setbacks.

Geoengineering research has a problem. That much should be clear following last week's cancellation of a field trial for the Stratospheric Particle Injection for Climate Engineering (SPICE) project. The solutions to this problem are not so obvious, but they must be found — and fast.

The SPICE field trial was supposed to involve spraying water into the atmosphere at an altitude of 1 kilometre using a balloon and hose-pipe, as part of a host of work exploring whether it is possible to mitigate global warming by introducing particles into the stratosphere to reflect some of the Sun's energy away from Earth.

But the field trial — which is only a small part of the overall SPICE project — became bogged down in protests and delays almost as soon as it was announced. Last week, as first reported by *Nature*, the project's lead investigator announced that it was being abandoned, citing concerns about intellectual-property rights, public engagement and the overall governance regime for such work.

Colleagues have leapt to the defence of the SPICE team, and praised its decision to continue with the theoretical strands of its work. Indeed, the researchers have acted with commendable honesty. But the SPICE issue is a perfect example of the problems that will persist until geoengineers grasp the nettle of regulation and oversight.

We have been here before. Work on 'fertilizing' the oceans to promote blooms of phytoplankton that would lock up carbon dioxide ran into similar protests and governance wrangles. In 2009, an experiment to test the idea by dumping tonnes of iron sulphate into the Southern Ocean caused huge public disquiet and went ahead only after further discussions.

Researchers argue that 'geoengineering' is a falsely inclusive term. They say that SPICE-style 'solar-radiation management' is completely different from ocean fertilization, and different again from carbon capture. But these technologies have similar aims and, when it comes to rules and regulations, they probably need to be dealt with together.

The geoengineering community has tried to bring some discipline to the emerging field. The 'Oxford Principles' — developed in 2010 by researchers at the University of Oxford, UK — offer some useful ground rules. They say that geoengineering should be regulated as a public good; there should be public participation in decision-making; research should be disclosed and results published openly; impacts should be assessed independently; and decisions to deploy the technologies should be made within a robust governance framework.

These are excellent principles. But they are vague, and cannot serve as a guide to conducting specific experiments in such a broad field.

A meeting of geoengineers in Asilomar, California, in 2009 — influenced by a meeting at the same location in 1975, when researchers hashed out guidelines for genetic engineering — produced similarly vague recommendations, such as the need to conduct research openly and to consult the public when planning research. It also called for

governments to "when necessary, create new mechanisms for the governance and oversight of large-scale climate engineering research activities".

The SPICE fiasco starkly demonstrates the need for such mechanisms. For a project of such high profile to founder on problems of

"Problems will persist until geoengineers grasp the nettle of regulation and oversight."

intellectual property, regulation or public protest would be bad enough. That it ran into difficulties in all three areas shows an underlying problem.

Of the issues raised, intellectual property may turn out to be the easiest to resolve (see page 429). Science has a long and generally happy relationship with patents, including those for technology with the ability to drive worldwide change. Likewise, lessons on public engagement and dealing with protests can be taken from earlier rows over genetic modification, stem cells, fertility work and animal research.

More troubling is the lack of an overarching governance framework. Although the SPICE trial has been cancelled, other tests of geoengineering technology will surely follow. Other work, such as fiddling with clouds to make them more reflective or to try to bring on rain, touches on both climate-change mitigation and weather modification.

Geoengineers should keep trying. They should come together and draft detailed, practical actions that need to be taken to advance governance in the field. Regulation in these cutting-edge and controversial areas needs to be working before the experiments begin, rather than racing to catch up. ■

In from the cold

Research into the Fukushima meltdowns has given a new lease of life to a valuable group.

The United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) is a relic of the cold war. Established in 1955 to study fallout from above-ground nuclear-weapons tests, the committee, which is based in Vienna, acted as one of the few channels for cooperation between the United States and the Soviet Union, and served to exchange information between East and West.

It was invaluable in its time: after the catastrophic meltdown at the Chernobyl nuclear power plant in Ukraine in 1986, the committee's ties to the Soviet Union allowed it to produce some of the first

independent assessments of the accident's aftermath and probable implications, at a time when they were sorely needed.

In recent years, however, UNSCEAR has seemed less useful. The cold war has ended, and above-ground nuclear testing is considered so gauche that even North Korea is unwilling to do it. The committee's weighty reports are no longer the first port of call for scientists studying the health effects of radiation, who prefer the convenience of online searches of the open literature. Even within UNSCEAR's small secretariat, some people wondered whether the committee had run its course.

And then, in 2011, came the meltdowns at the Fukushima Daiichi nuclear plant in Japan. Three reactors released radioactive contaminants into the environment and forced the evacuation of thousands of people. Phones at UNSCEAR started to ring again.

Asked to investigate, the committee brought together dozens of scientists from universities and government agencies around the world to review the available data. They are now mid-way through their work, and much remains to be done, but already the value of the exercise, and of the organization behind it, is becoming clear.

In the days after the Fukushima accident, Japanese citizens were bombarded with radiation readings. The numbers were delivered in obscure units with little or no context, along with limp reassurances from government authorities. The data were inconsistent, and frequently wrong. Those early readings have been followed by government surveys, academic assessments and independent models of varying quality and message. The cacophony has fed mistrust and fear among local residents and the wider public.

UNSCEAR, however, is carefully reviewing the shaky readings taken during the early days of the accident to establish which of them are useful. It is integrating data from a wide range of sources to see what information has been collected and how it has been calibrated. And it is building its own models to reconstruct exactly what was emitted by the reactors and when. The aim is to come up with a coherent picture of the accident by the end of next year. Much remains to be done, but

this week UNSCEAR's working committee on Fukushima has been able to provide a comprehensive — and seemingly reassuring — view of radiation exposure among workers at the plant. That, combined with the available data on public exposure, indicates to many experts that the health effects from the accident will be minimal (see page 423).

Although some of the committee's scientists believe that its work will provide a measure of reassurance to the people of Japan, that is by no

“UNSCEAR will provide a strong, scientifically sound account of the Fukushima accident.”

means certain. Many of UNSCEAR's members come from regulatory bodies in their home nations — what critics of the nuclear industry view as the establishment. As such, they may struggle to convince outsiders that they act in good faith. One way to build support and bridges with potential critics might be to take more seriously the data collected by

independent groups such as concerned citizens and international non-governmental organizations in the immediate aftermath of the disaster. The committee is already considering this, and it should proceed.

There will remain limits to the comfort that science can provide. Cancer is a disease of statistics: some workers will get ill and some will blame it on the dose they received. Indeed, one has already died of leukaemia that is, as far as anyone can tell, unrelated to the accident.

Nevertheless, UNSCEAR has been reborn. The committee is no longer a clearing house — it is a filter. Its work in the coming months and years will provide a strong, scientifically sound account of the Fukushima accident. It will bring consistency to the numbers and mould a single narrative from the disparate data sets. Not everyone will agree with the committee's conclusions, but they will provide a definitive point of reference for discussion and evaluation.

Accidents such as that at Fukushima are thankfully rare, and within a few years UNSCEAR may fade back into obscurity. It can be allowed to slumber, but it must not be forgotten. In the modern information age, its purpose is clearer than ever before. ■

A whale of a story

A previously unknown sensory organ provides a lesson in coordination.

“To produce a mighty book, you must choose a mighty theme. No great and enduring volume can ever be written on the flea, though many there be that have tried it.” Herman Melville, the author of *Moby Dick* and of that quote, certainly knew a mighty theme when he saw one. His masterpiece is as much a study on the behaviours of the sperm whales that Melville saw while serving on a whaling ship as it is an exploration of the motives and characters of the men who pursue them.

Sperm whales (*Physeter catodon*) have teeth, as Melville and his fictional Captain Ahab knew all too well. So how they catch and eat their prey, although a mighty theme, is not much of a story. More interesting, and much more mysterious, is how rorqual whales — those toothless giants of the oceans — manage to synchronize the multiple complex processes that they use to survive on little more than krill and small fish.

These creatures, which include the fin (*Balaenoptera physalus*), humpback (*Megaptera novaeangliae*) and blue whale (*Balaenoptera musculus*), have long puzzled biologists. How can they survive and sustain their huge bodies on such a meagre diet? How can they get enough food? The answer involves one of the greatest feats in the animal kingdom.

Rorqual whales capture much of their food by an extraordinary procedure known as lunge feeding. When a rorqual comes across a dense patch of prey, it accelerates through the water and open its mouth. As it

does so, its mouth fills with water, suspended within which are the tiny animals that the whale wants. The amount of water that flows into the whale can more than double the creature's weight, and to accommodate it, blubbery pleats under the lower jaw expand, just as an accordion grows as it fills with air. The once sleek and streamlined whale now has the shape of a bloated tadpole. And it has a lot of water in its mouth.

To squeeze the water out again, the whale closes its jaws and pushes the water out through plates of keratin filters, which trap the food. In this way, rorqual whales can gulp and graze for hours, repeatedly slowing down then lunging through the water.

It is a unique process, and one that requires some special equipment. The shape of a rorqual's head is more reptilian than mammalian, with a giant skull and mandibles that are held together by unique jaw joints made from dense elastic matrices of fibre and cartilage and infused with oil. This system lets the whale open its jaws to nearly a right angle — ideal for gulping water. There is more: the lower mandible has a flexible joint at its centre, which allows the two sides to rotate. And the whale can turn its tongue inside out, pulling it backwards to create even more space for the water. As it does so, its open mouth experiences massive drag, which the creature mitigates by contracting its throat to force some of the water back out.

But how on Earth does a rorqual manage to coordinate this activity? On page 498, Nicholas Pyenson of the Smithsonian Institution in Washington DC and his colleagues describe a sensory organ that they discovered in the jaws of several species that might offer an answer. The organ, they say, seems to be involved in several lunge-feeding processes, including the opening of the jaws and the expansion of the throat. And that means that it

➔ **NATURE.COM**

To comment online,
click on Editorials at:
go.nature.com/xhunjv

may have helped to drive the evolution of the huge bodies of rorqual whales — their great and enduring volumes, as Melville might have put it. ■

independent assessments of the accident's aftermath and probable implications, at a time when they were sorely needed.

In recent years, however, UNSCEAR has seemed less useful. The cold war has ended, and above-ground nuclear testing is considered so gauche that even North Korea is unwilling to do it. The committee's weighty reports are no longer the first port of call for scientists studying the health effects of radiation, who prefer the convenience of online searches of the open literature. Even within UNSCEAR's small secretariat, some people wondered whether the committee had run its course.

And then, in 2011, came the meltdowns at the Fukushima Daiichi nuclear plant in Japan. Three reactors released radioactive contaminants into the environment and forced the evacuation of thousands of people. Phones at UNSCEAR started to ring again.

Asked to investigate, the committee brought together dozens of scientists from universities and government agencies around the world to review the available data. They are now mid-way through their work, and much remains to be done, but already the value of the exercise, and of the organization behind it, is becoming clear.

In the days after the Fukushima accident, Japanese citizens were bombarded with radiation readings. The numbers were delivered in obscure units with little or no context, along with limp reassurances from government authorities. The data were inconsistent, and frequently wrong. Those early readings have been followed by government surveys, academic assessments and independent models of varying quality and message. The cacophony has fed mistrust and fear among local residents and the wider public.

UNSCEAR, however, is carefully reviewing the shaky readings taken during the early days of the accident to establish which of them are useful. It is integrating data from a wide range of sources to see what information has been collected and how it has been calibrated. And it is building its own models to reconstruct exactly what was emitted by the reactors and when. The aim is to come up with a coherent picture of the accident by the end of next year. Much remains to be done, but

this week UNSCEAR's working committee on Fukushima has been able to provide a comprehensive — and seemingly reassuring — view of radiation exposure among workers at the plant. That, combined with the available data on public exposure, indicates to many experts that the health effects from the accident will be minimal (see page 423).

Although some of the committee's scientists believe that its work will provide a measure of reassurance to the people of Japan, that is by no

“UNSCEAR will provide a strong, scientifically sound account of the Fukushima accident.”

means certain. Many of UNSCEAR's members come from regulatory bodies in their home nations — what critics of the nuclear industry view as the establishment. As such, they may struggle to convince outsiders that they act in good faith. One way to build support and bridges with potential critics might be to take more seriously the data collected by

independent groups such as concerned citizens and international non-governmental organizations in the immediate aftermath of the disaster. The committee is already considering this, and it should proceed.

There will remain limits to the comfort that science can provide. Cancer is a disease of statistics: some workers will get ill and some will blame it on the dose they received. Indeed, one has already died of leukaemia that is, as far as anyone can tell, unrelated to the accident.

Nevertheless, UNSCEAR has been reborn. The committee is no longer a clearing house — it is a filter. Its work in the coming months and years will provide a strong, scientifically sound account of the Fukushima accident. It will bring consistency to the numbers and mould a single narrative from the disparate data sets. Not everyone will agree with the committee's conclusions, but they will provide a definitive point of reference for discussion and evaluation.

Accidents such as that at Fukushima are thankfully rare, and within a few years UNSCEAR may fade back into obscurity. It can be allowed to slumber, but it must not be forgotten. In the modern information age, its purpose is clearer than ever before. ■

A whale of a story

A previously unknown sensory organ provides a lesson in coordination.

“To produce a mighty book, you must choose a mighty theme. No great and enduring volume can ever be written on the flea, though many there be that have tried it.” Herman Melville, the author of *Moby Dick* and of that quote, certainly knew a mighty theme when he saw one. His masterpiece is as much a study on the behaviours of the sperm whales that Melville saw while serving on a whaling ship as it is an exploration of the motives and characters of the men who pursue them.

Sperm whales (*Physeter catodon*) have teeth, as Melville and his fictional Captain Ahab knew all too well. So how they catch and eat their prey, although a mighty theme, is not much of a story. More interesting, and much more mysterious, is how rorqual whales — those toothless giants of the oceans — manage to synchronize the multiple complex processes that they use to survive on little more than krill and small fish.

These creatures, which include the fin (*Balaenoptera physalus*), humpback (*Megaptera novaeangliae*) and blue whale (*Balaenoptera musculus*), have long puzzled biologists. How can they survive and sustain their huge bodies on such a meagre diet? How can they get enough food? The answer involves one of the greatest feats in the animal kingdom.

Rorqual whales capture much of their food by an extraordinary procedure known as lunge feeding. When a rorqual comes across a dense patch of prey, it accelerates through the water and open its mouth. As it

does so, its mouth fills with water, suspended within which are the tiny animals that the whale wants. The amount of water that flows into the whale can more than double the creature's weight, and to accommodate it, blubbery pleats under the lower jaw expand, just as an accordion grows as it fills with air. The once sleek and streamlined whale now has the shape of a bloated tadpole. And it has a lot of water in its mouth.

To squeeze the water out again, the whale closes its jaws and pushes the water out through plates of keratin filters, which trap the food. In this way, rorqual whales can gulp and graze for hours, repeatedly slowing down then lunging through the water.

It is a unique process, and one that requires some special equipment. The shape of a rorqual's head is more reptilian than mammalian, with a giant skull and mandibles that are held together by unique jaw joints made from dense elastic matrices of fibre and cartilage and infused with oil. This system lets the whale open its jaws to nearly a right angle — ideal for gulping water. There is more: the lower mandible has a flexible joint at its centre, which allows the two sides to rotate. And the whale can turn its tongue inside out, pulling it backwards to create even more space for the water. As it does so, its open mouth experiences massive drag, which the creature mitigates by contracting its throat to force some of the water back out.

But how on Earth does a rorqual manage to coordinate this activity? On page 498, Nicholas Pyenson of the Smithsonian Institution in Washington DC and his colleagues describe a sensory organ that they discovered in the jaws of several species that might offer an answer. The organ, they say, seems to be involved in several lunge-feeding processes, including the opening of the jaws and the expansion of the throat. And that means that it

➔ **NATURE.COM**

To comment online,
click on Editorials at:
go.nature.com/xhunjv

may have helped to drive the evolution of the huge bodies of rorqual whales — their great and enduring volumes, as Melville might have put it. ■



We must set planetary boundaries wisely

The concept of environmental thresholds is compelling, but it has the potential to shift political focus to the wrong areas, says **Simon L. Lewis**.

As pressure on resources increases, pollution accumulates and humanity's impact on Earth escalates, global-scale governance of the environment is increasingly necessary. In June, the United Nations' Rio+20 Earth Summit in Rio de Janeiro, Brazil, will grapple with these difficult political issues. Up for discussion is a relatively new scientific concept: planetary boundaries.

Formulated in 2009 by Johan Rockström, director of the Stockholm Environment Institute, and his colleagues, the concept is based on the idea that humanity flourished under the conditions on Earth in the 10,000 years leading up to the industrial revolution — the Holocene epoch. So, to maintain human progress, we should keep the planet under similar biophysical conditions. The researchers set out nine key environmental measures and thresholds that should not be breached for fear of pushing Earth out of the Holocene-like 'safe operating space for humanity'. The boundaries include thresholds for climate change and biodiversity loss that have already been crossed.

The idea is conceptually brilliant and politically seductive: clear, quantitative measurements with no obvious judgements on what is 'right' or 'wrong' to include. It is also liberating. Here is humanity's safe space: within it, do what you want.

In Rio, there will be a push to convert the concept into meaningful action. It has support from UN secretary-general Ban Ki-moon, and is included in the conference's draft negotiating text. Yet the boundaries concept has two important flaws, and using it uncritically could unwittingly undermine Rio's twin goals of environmental stewardship and ensuring a good life for everyone.

The first flaw, from a human-welfare perspective, is that not all of the identified parameters are true thresholds that, once passed, can be recovered to move back to Holocene-like conditions. Some parameters are fixed limits, not boundaries. Take disruption of the phosphorus cycle: this is represented in the planetary boundaries concept as the quantity of phosphates flowing into the oceans from crop-fertilizer run-off, which can cause algal blooms and an oxygen deficit for marine life. Framed in this way — 'don't destroy the marine environment' — the boundary makes sense. But more serious for humanity is that phosphorus is a key plant nutrient. Fertilizer is produced from rock phosphate, which forms on geological time scales. When it is gone, it is gone. This does not represent a threshold boundary: it is a depletion-limit. Humanity cannot use more rock phosphate than there is.

This distinction between thresholds (which we can breach), and fixed limits (which we cannot) may seem academic, but it has important policy implications. To highlight a boundary on phosphate pollution, for example, would drive investment in technology to combat the impact on marine

environments, but do nothing to stop the running down of rock-phosphate supplies. To emphasize the depletion limit would shift the focus to technology to use and re-use phosphorus to safeguard stocks.

Similarly, at the Planet Under Pressure conference in London in March, US scientist Steven Running proposed a new planetary boundary: terrestrial net primary productivity (NPP), or more simply, plant growth. Despite massive agricultural expansion in the past century, global NPP has not dramatically increased. It is a ceiling limit. Thus, the allocation of NPP to benefit biodiversity or food, fibre, fodder and fuel for humans is essentially a zero-sum activity. Yet Rockström's published planetary boundary suggests that we could expand croplands by 400 million hectares before reaching the threshold — something that would seriously harm biodiversity.

The second weakness relates to scale. True threshold boundaries come in two types. Some are unambiguously global, such as climate change, which is driven by well-mixed greenhouse gases in the atmosphere. Others, such as nitrogen pollution, are global only if local problems are widely replicated. Even if nitrogen-fertilizer run-off is cleaned up in China's Yangtze River, it has no direct impact on nitrogen pollution in Nigeria's Niger Delta. These are regional problems, but in aggregate can be of global significance. The planetary boundaries concept does note that whereas climate change, ocean acidification and stratospheric ozone depletion are 'systemic processes', the rest are 'aggregate processes'. However, each published safe threshold is based on a single global number, and will probably be treated accordingly.

A global focus on nine boundaries could spread political will thinly — and it is already weak. There is no need for all the world's countries to enter protracted legal discussions on aggregate boundaries: those affected by regional problems should work among themselves to solve them. Global negotiations should focus on managing the clear global planetary boundaries of climate change and ocean acidification, as well as biodiversity loss, which has global drivers.

The concept of planetary boundaries and avoiding dangerous thresholds is important but limited. Furthermore, a narrow focus on maintaining Holocene-like conditions risks side-lining key problems such as the 'plastic soup' of particulate waste that stretches across the Pacific Ocean. This does not fit the boundaries model, because there was no plastic during the pre-industrial Holocene. A simple transfer of a neat scientific idea into the policy arena could cause as many problems for policy-makers as it solves. ■

Simon L. Lewis is reader in global change science at University College London and the University of Leeds, UK.
e-mail: s.l.lewis@leeds.ac.uk

**BOUNDARIES COULD
SPREAD
POLITICAL
WILL
THINLY — AND IT IS
ALREADY
WEAK.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/4ukvxx

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

HUMAN GENETICS

Not-so-rare gene variants

By sequencing more human genomes more thoroughly than ever before, researchers have shown that most of the genetic variation in the human population is due to rare variants. These rare versions of genes are more likely to affect human health than are their common counterparts — which could make it more difficult to link gene variants to disease than previously thought.

Joshua Akey and Michael Bamshad at the University of Washington, Seattle, and their colleagues sequenced 15,585 genes in 2,440 people. The researchers found that 86% of the gene variants that they detected were rare, as were more than 95% of variants that were predicted to have a medical or biological effect.

Meanwhile, a team including Matthew Nelson at GlaxoSmithKline in Research Triangle Park, North Carolina, and John Novembre at the University of California, Los Angeles, report that more than 74% of variants identified by sequencing 202 genes in 14,002 people were carried by only one or two of the individuals.

Andy Clark and Alon Keinan of Cornell University in Ithaca, New York, used data from previous sequencing studies to show that rapid growth of the human population, starting 1,400 years ago, helped to seed the human genome with numerous rare variants.

Science <http://dx.doi.org/10.1126/science.1219240>; <http://dx.doi.org/10.1126/science.1217876>; 336, 740–743 (2012)

For a longer story on this research, see go.nature.com/lssahs



HYDROLOGY

Groundwater down, sea level up

Groundwater depletion and other human-induced changes in terrestrial water storage were responsible for almost half of the increase in global sea level observed between 1961 and 2003.

Thermal expansion of the oceans, the melting of polar ice and changes in terrestrial water reservoirs all contribute to rising sea levels. To single out the impact of changes in land-based water use on sea level, Yadu Pokhrel at the University of Tokyo and his colleagues ran a simulation of global terrestrial water stocks and flows, accounting for human activities such as irrigation (pictured), which often uses

groundwater, and dam-building.

The model suggests that about 42% of the almost eight-centimetre rise in sea level observed over the study period has resulted from changes in terrestrial water storage, particularly groundwater use. This human contribution could explain the discrepancy between observed and expected sea levels, the authors say.

Nature Geosci. <http://dx.doi.org/10.1038/ngeo1476> (2012)

For a longer story on this research, see go.nature.com/damx2m

NEUROSCIENCE

Brain stimulation zaps fear

Electrical stimulation of certain brain regions has been used to treat obsessive-compulsive disorder (OCD), but how it works is unknown. Researchers suggest that the stimulation could work by acting on brain circuits involved in dampening fear responses.

Gregory Quirk and his team at the University of Puerto Rico in San Juan implanted electrodes in the

brains of rats. The authors conditioned the animals to expect a footshock whenever they heard a tone and then, the next day, exposed the rats to the tone without the shock to extinguish the conditioned fear. When, on the second day, the researchers stimulated a specific area of a brain region similar to that targeted during OCD treatment, the rats later showed less fear on hearing the tone than did control rats. The stimulation also boosted the expression of a marker of plasticity — the ability to reorganize neural

pathways — in regions involved in extinguishing fear. This suggests that the brain stimulation enhances fear extinction.

Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1200782109> (2012)

BIOENGINEERING

Viruses as power generators

Thin films of viruses can generate an electrical charge when subjected to mechanical stress — a form of energy

conversion known as the piezoelectric effect.

Seung-Wuk Lee and his team at the Lawrence Berkeley National Laboratory, California, measured the charge generated by stressing self-assembled films of M13 bacteriophage, a virus that attacks only bacteria. The authors boosted the films' power output by modifying the proteins that form the phage outer coat and by stacking the films on top of each other. The team went on to build a penny-sized generator containing the films, which produced enough power to light up a liquid-crystal display.

The viral power generator could be used as an environmentally friendly component in nanodevices, the authors suggest.

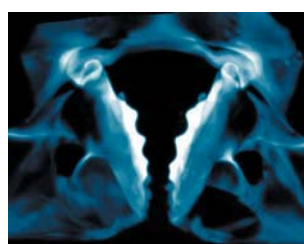
Nature Nanotechnol. <http://dx.doi.org/10.1038/nnano.2012.69> (2012)

EVOLUTION

Tooth-enamel similarities

The teeth of a freshwater crustacean are coated with a wear-resistant, enamel-like layer similar to that on vertebrate teeth. This is an example of convergent evolution in which a biological trait develops independently in separate lineages.

Researchers including Amir Berman at Ben-Gurion University in Beer-Sheva, Israel, and Barbara Aichmayer at the Max Planck Institute of Colloids and Interfaces in Potsdam, Germany, used Raman spectroscopy to analyse the mandibles of *Cherax quadricarinatus* crayfish. On the surface of the molar teeth, the researchers



found a layer of fluorapatite crystals (**pictured, white**), which are chemically similar to the carbonated hydroxyapatite found in vertebrate teeth.

The results indicate that comparable chewing needs led to the independent evolution of similar tooth structures in species from different lineages. *Nature Commun.* **3**, 839 (2012)

NUCLEAR PHYSICS

Fingerprint of nuclear fuel

Plutonium-239 is used in nuclear weapons, but it has been difficult to detect in the lab. Researchers in New Mexico have now succeeded in measuring the isotope's nuclear magnetic resonance (NMR) signal.

Interactions within plutonium atoms have frustrated attempts over the past 50 years to find the NMR signal of pure plutonium. To reduce the impact of these interactions, Georgios Koutroulakis and his colleagues at the Los Alamos National Laboratory analysed plutonium dioxide, PuO₂ — which has a more stable NMR signal than other plutonium compounds — at a very low temperature.

Teasing out the atom's signature in other compounds could be difficult, but at least now researchers know where to look, the authors say.

Science **336**, 901–904 (2012)

ECOLOGY

Outpaced by climate change

Mammals could find it more difficult to adjust their geographical ranges as the climate warms than previously thought.

Carrie Schloss and her colleagues at the University of Washington, Seattle, studied the ability of 493 mammalian species to move to new habitats with suitable climates when the projected temperature increases in the

COMMUNITY CHOICE

The most viewed papers in science

GEOENGINEERING

Cooling effects of white roofs

HIGHLY READ
on *iopscience*.
iop.org
15 Apr–15 May

Increasing the ability of urban roofs and pavements to reflect the Sun's heat would decrease the average global temperature by less than 0.1 °C over the next 300 years, according to climate simulations.

Hashem Akbari and his colleagues at Concordia University in Montreal, Canada, used a global climate model, along with estimates of the extent of the world's urban areas, to predict what would happen if urban solar reflectivity rose from roughly 15% to 25%. Depending on which set of urban data they used, the researchers found that the resulting global cooling ranged from 0.01 to 0.07 °C by 2300.

Although this temperature decrease is small relative to natural climate variability, the authors say that cities should promote the use of white or light-coloured surface materials because these can reduce air-conditioning usage and counteract the urban 'heat island' effect at minimal extra cost.

Environ. Res. Lett. **7**, 024004 (2012)

Western hemisphere occur. The team modelled the speed at which species would need to move and compared this with their known dispersal rates. The authors found that, on average, 9.2% of species at any given location will be unable to outrun climate change. Reductions in range size will be widespread, affecting 87%, with 20% of these reductions due to species' inability to relocate. Primates, which do not disperse as readily as other mammals, are likely to be among the most vulnerable.

Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1116791109> (2012)

ZOOLOGY

Star's five-legged coordination

Brittle stars use their five limbs to crawl along the sea floor in a bilateral manner, even though their bodies have radial, not bilateral, symmetry.

Henry Astley at Brown University in Providence, Rhode Island, measured the movements of 13 *Ophiocoma echinata* brittle stars



(**pictured**). He found that the creatures coordinate their limb movements, rather than having each leg react independently to stimuli. The animals move along the axis of a central limb, held either in front of or behind the body, and turn by re-designating the central limb rather than by rotating the body.

Brittle stars are unusual in having limbs with different, interchangeable roles in locomotion that are not constrained by anatomy. *J. Exp. Biol.* **215**, 1923–1929 (2012)

➔ **NATURE.COM**

For the latest research published by Nature visit:
www.nature.com/latestresearch

H. ASTLEY

SEVEN DAYS

The news in brief

POLICY

South Korea vision

South Korea's government officially launched its ambitious Institute for Basic Science (IBS) on 17 May. The government hopes that the institute, headquartered in Daejeon, will eventually rank alongside Germany's Max Planck Society and Japan's RIKEN as a world-leading institution. The IBS plans to attract 3,000 researchers and staff members to 50 research centres around the country, each with an average annual budget of 10 billion won (US\$8.6 million). Along with the construction of a rare-isotope accelerator, the IBS is one of the main initiatives of South Korea's 5.2-trillion-won 'science belt' project. See go.nature.com/uhjve1 for more.

SPICE trial stopped

UK-based researchers have called off an experiment to test climate-engineering technology by spraying water from a balloon 1 kilometre above Earth, part of the Stratospheric Particle Injection for Climate Engineering (SPICE) project. The lack of rules governing such geoengineering experiments, a potential conflict of interest over a patent application for some of the technology involved and the need for more public engagement all contributed to the decision, said principal investigator Matthew Watson, an Earth scientist at the University of Bristol, UK. See pages 415 and 429 for more.

Global research

The heads of around 50 national funding agencies took part in the inaugural meeting of the Global Research Council (GRC) at the headquarters of the US National Science Foundation



FISHERIES AND OCEANS CANADA

Canada cuts freshwater research

Canada's government has cancelled its funding for the Experimental Lakes Area (ELA), a site in Ontario where influential field research has been done for more than half a century. Fisheries and Oceans Canada told its staff on 17 May that research at the ELA, a collection of 58 lakes and a laboratory complex, would be shut down in March 2013. Studies at the

site (pictured) have influenced international policies, most notably a 1991 air-quality agreement between the United States and Canada, which led to reductions in acid rain. Ecologists condemned the decision, the latest in a string of moves to scale back or shut down research programmes in the wake of budget cuts. See go.nature.com/5fxqd2 for more.

in Arlington, Virginia, on 13–15 May. The GRC, a forum that will examine issues important to science-funding agencies worldwide, produced a set of international principles on peer review; it will next discuss open access and research integrity. See page 427 for more.

Coffin protest

Around 80 physicists and chemists in the United Kingdom have staged an unusual protest against their research funder: carrying a coffin down London's Whitehall to Downing Street. The stunt, performed on 15 May, marks the start of a campaign pressing for reform of the Engineering and Physical Sciences Research Council (EPSRC). Protesters

say that the council's policies have resulted in money being allocated to strategic priorities rather than to blue-skies research; the EPSRC said that the protesters were a small, vocal minority of scientists. See go.nature.com/demha6 for more.

Chilean dams

The Chilean Supreme Court has blocked the construction of a large hydroelectric dam in Patagonia, overturning the decision of a regional review board that had approved the project despite opposition from environmental groups. The 640-megawatt Rio Cuervo dam, developed by the Australian firms Origin Energy of Sydney and Xstrata Copper in Brisbane, is part of a three-dam system that

would generate one gigawatt of power. On 11 May, the court ruled that project backers had not completed an essential geological survey. The court has, however, already given the green light to the construction of the larger five-dam, 2.75-gigawatt HidroAysén project in the region.

Prostate screening

A US government advisory task force has recommended against the routine use of a controversial test for prostate cancer to screen healthy men, because its harms outweigh its benefits. The final recommendation from the US Preventive Services Task Force, published on 21 May (V. A. Moyer *et al.* *Ann. Intern. Med.* advanced online publication, 2012;

see go.nature.com/xjusen), immediately drew a backlash from physicians, who said that some patients would die from prostate cancer if routine screening for prostate-specific antigen (PSA) were stopped. The issue, which pits public-health evidence against politics, is contested in the United States; in other countries, the PSA test is less widely used (see *Nature* **478**, 286; 2011).

PEOPLE

Nuclear stand-down

The embattled chairman of the US Nuclear Regulatory Commission, Gregory Jaczko, announced his resignation on 21 May, potentially signalling an end to three years of controversy, acrimony and stalemate over US nuclear-waste policy. Jaczko's leadership has been beset by bitter policy disputes over the closure of the Yucca Mountain nuclear-waste repository in Nevada, as well as accusations that his management style created a hostile work environment for staff members. See go.nature.com/wbyjjq for more.

French science head

Geneviève Fioraso (**pictured**) has been appointed minister of higher education and research in France's new government. Fioraso, a socialist parliamentarian and the deputy mayor of



Grenoble, has experience in research and innovation. Starting out as a lecturer in English and economics, she has managed high-tech start-up firms, authored a report on synthetic biology for France's Parliamentary Office for the Evaluation of Scientific and Technological Choices, and is currently chief executive of the business-incubator wing of Minatéc — Grenoble's innovation campus for nanotechnology and electronics. See go.nature.com/kayiqy for more.

New head for MIT

Rafael Reif, an electrical engineer who has been provost of the Massachusetts Institute of Technology (MIT) in Cambridge since 2005, will be the institution's next president. MIT announced Reif's election on 16 May; he will take up the post on 2 July. Reif replaces neuroscientist Susan Hockfield, who is standing down after seven years in charge.

EVENTS

GM crop attack

A trial crop of genetically modified (GM) wheat at Britain's leading agricultural research centre, Rothamsted Research in Harpenden, was attacked on 20 May. The institute said that a lone protester broke in and damaged property, but that the trial was unaffected. He has since been charged by police. A protest group called 'Take the Flour Back', which has been publicizing its intention to destroy the crop on 27 May, says that its action will go ahead as planned. See go.nature.com/zdvne for more.

Gas leak plugged

A major gas leak at a drilling platform in the Elgin gas field in the North Sea has been stopped after 52 days. The site's owner, oil company Total, pumped heavy mud into a well to plug the leak, and announced on 16 May that the effort had been a success. The leak, which began on 25 March, had triggered an evacuation of the site and was one of the highest-profile drilling accidents since the Gulf of Mexico Deepwater Horizon disaster in 2010.

Private space flight

The first commercial cargo mission to the International Space Station (ISS) — a

COMING UP

29 MAY

In London, the International Energy Agency issues a report on the 'golden age of gas', evaluating the potential of shale-gas resources found in recent years, and concerns about how to exploit them.

30–31 MAY

Research ministers from European Union (EU) member states meet in Brussels to discuss detailed proposals for Horizon 2020, the EU's €80-billion (US\$102-billion) research-funding programme for 2014–20.

go.nature.com/8vn3om

milestone for private space flight — lifted off from Cape Canaveral in Florida on 22 May. Launched by SpaceX of Hawthorne, California, the Falcon 9 rocket carried the company's Dragon capsule, packed with some 460 kilograms of supplies for the ISS crew. The mission aim is for the capsule to dock with the ISS on 25 May and spend 18 days attached to the station before splashing down in the Pacific Ocean. See go.nature.com/hrg5mh for more.

RESEARCH

Alzheimer's plan

The US government has released its first strategic plan for battling Alzheimer's disease — the progressive, incurable dementia that afflicts more than 5 million Americans. The plan, released on 15 May, was mostly unchanged from a draft released in February, and aims to find ways to treat and prevent Alzheimer's by 2025. See page 426 for more.

► **NATURE.COM**

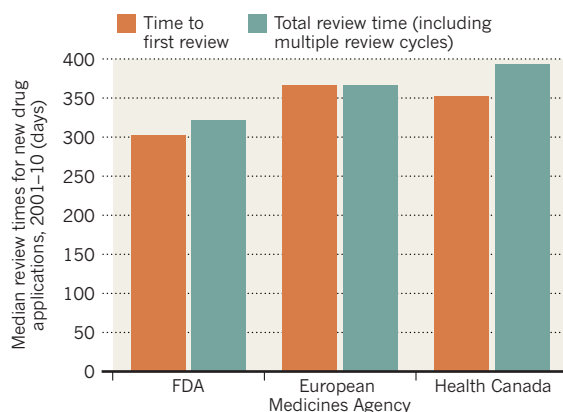
For daily news updates see:
www.nature.com/news

TREND WATCH

The US Food and Drug Administration is often criticized for taking a plodding approach to drug regulation. But when it comes to approving novel therapeutics, the agency is actually quicker than its European and Canadian counterparts. Researchers from Yale University in New Haven, Connecticut, reported the findings on 16 May (N. S. Downing *et al.* *N. Engl. J. Med.* <http://dx.doi.org/10.1056/NEJMsa1200223>; 2012). See go.nature.com/brlmhf for more.

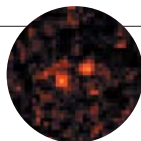
SPEEDY DRUG REVIEWS

The US Food and Drug Administration (FDA) is quicker to review new drugs than its European and Canadian counterparts.



NEWS IN FOCUS

ASTRONOMY Decision time for disputed planet claim **p.428**



CLIMATE CHANGE Patent row deflates balloon-borne experiment **p.429**

AGRICULTURE New strategies to control herbicide-resistant superweeds **p.430**

PHYSICS Italy's palace for particle physics, 1,400 metres underground **p.435**

Y. TSUNO/EPA/CORBIS



Around 170 of Fukushima's workers have a slightly elevated risk of cancer due to their radiation exposure.

RADIATION HEALTH

Fukushima's doses tallied

Studies indicate minimal health risks from radiation in the aftermath of Japan's nuclear disaster.

BY GEOFF BRUMFIEL

Few people will develop cancer as a consequence of being exposed to the radioactive material that spewed from Japan's Fukushima Daiichi nuclear power plant last year — and those who do will never know for sure what caused their disease. These conclusions are based on two comprehensive, independent assessments of the radiation doses received by Japanese citizens, as well as by the thousands of workers who battled to bring the shattered nuclear reactors under control.

The first report, seen exclusively by *Nature*, was produced by a subcommittee of the United

Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) in Vienna, and covers a wide swathe of issues related to all aspects of the accident. The second, a draft of which has been seen by *Nature*, comes from the World Health Organization (WHO) in Geneva, Switzerland, and estimates doses received by the general public in the first year after the accident. Both reports will be discussed at UNSCEAR's annual meeting in Vienna this week.

The UNSCEAR committee's analyses show that 167 workers at the plant received radiation doses that slightly raise their risk of developing cancer. The general public was largely protected by being promptly evacuated, although

the WHO report does find that some civilians' exposure exceeded the government's guidelines. "If there's a health risk, it's with the highly exposed workers," says Wolfgang Weiss, the chair of UNSCEAR. Even for these workers, future cancers may never be directly tied to the accident, owing to the small number of people involved and the high background rates of cancer in developed countries such as Japan.

Scientists involved in producing the UNSCEAR report hope that their independent summary of the best available data could help to dispel some of the fear about fallout that has grown over the past year (see *Nature* **483**, 138–140; 2012). As well as providing a preliminary assessment of workers' exposure, the UNSCEAR report concludes that the Japanese government's estimate of the radiation released was correct to within a factor of ten, and that further study is needed to fully understand the impacts of the accident on plants, animals and marine life near the power station. When a final version of the report is approved by the full UNSCEAR committee next year, it should provide a useful baseline for future studies.

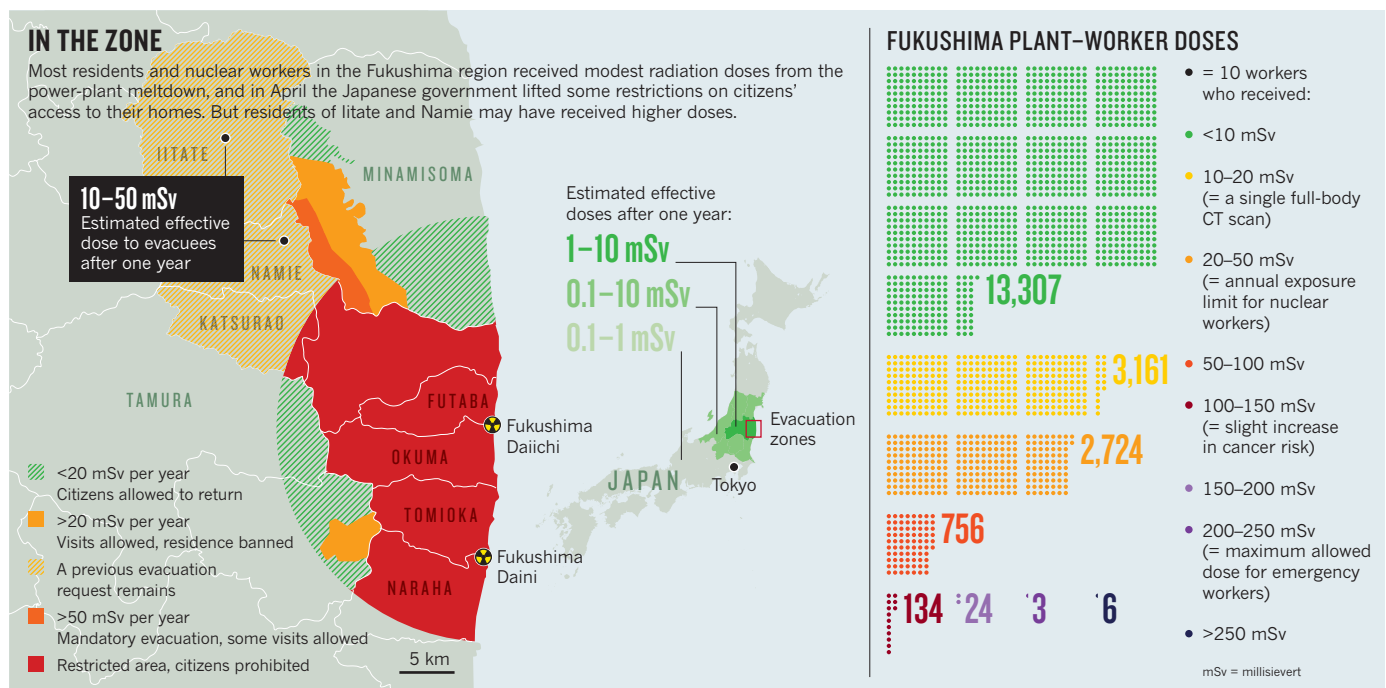
The Fukushima crisis began on 11 March 2011, when a magnitude-9.0 earthquake triggered a tsunami off the coast of Japan. A 14-metre wave flooded four of the six reactors at the Fukushima Daiichi plant, knocking out emergency cooling systems and leading to meltdowns and explosions that released radioactivity into the air and ocean. In the year since the accident, the plant has been stabilized, and radioactive emissions have largely stopped.

From last autumn, UNSCEAR has been reviewing all the available data on Fukushima's radiation — just as it did to produce what was then the definitive report on the 1986 Chernobyl nuclear accident. In particular, it scoured anonymized medical data for 20,115 workers and contractors employed by the Tokyo Electric Power Company, which runs the plant. It found that 146 employees and 21 contractors received a dose of more than 100 millisieverts (mSv), the level at which there is an acknowledged slight increase in cancer risk.

Six workers received more than the 250 mSv allowed by Japanese law for front-line emergency workers, and two operators in the control rooms for reactor units 3 and 4 received doses above ▶



WWW.NATURE.COM/JAPANQUAKE



► 600 mSv, because they had not taken potassium iodide tablets to help prevent their bodies from absorbing radioactive iodine-131 (see 'In the zone'). So far, neither operator seems to have suffered ill effects as a result of their exposure.

Most of the workers who received high doses were exposed in the early days of the crisis. In those first hours, they were huddled in darkened control rooms, while small teams made forays inside the reactor buildings to survey the damage and manually operate valves and other equipment. Often, they did not know how much radiation was present — the report says that an automated system designed to monitor their radiation levels was not operating properly. By mid-April, basic access control and monitoring had been restored on the site.

Experts agree that there is unlikely to be a detectable rise in thyroid cancer or leukaemia, the two cancers most likely to result from the accident. "There may be some increase in cancer risk that may not be detectable statistically," says Kiyohiko Mabuchi, who heads Chernobyl studies at the National Cancer Institute in Rockville, Maryland. In Chernobyl, where clean-up workers were exposed to much higher doses, 0.1% of the 110,000 workers surveyed have so far developed leukaemia, although not all of those cases resulted from the accident.

The risk to the roughly 140,000 civilians who had been living within a few tens of kilometres of the plant seems even lower. Because detailed radiation measurements were unavailable at the time of the accident, the WHO estimated doses to the public, including radiation exposure from inhalation, ingestion and fallout. The agency concludes that most residents of Fukushima and neighbouring Japanese prefectures received a dose below 10 mSv. Residents

of Namie town and Iitate village, two areas that were not evacuated until months after the accident, received 10–50 mSv. The government aims to keep public exposure from the accident below 20 mSv per year, but in the longer term it wants to decontaminate the region so that residents will receive no more than 1 mSv per year from the accident.

The WHO's calculations are consistent with several health surveys conducted by Japanese scientists, which found civilian doses at or below the 1–15-mSv range, even among people living near the plant. One worrying exception is that infants in Namie town may have been exposed to enough iodine-131 to receive an estimated thyroid dose of 100–200 mSv, raising their risk of thyroid cancer. But data collected from 1,080 children in the region found that none had received a thyroid dose greater than 50 mSv. Chernobyl's main cancer legacy in children was thyroid cancer.

FEARFUL AND ANGRY

The large population involved could mean that the eventual number of radiation-induced cancers among the public will actually be higher than among workers, even though the risk to each individual civilian is tiny, says David Brenner, a radiologist at Columbia University in New York city. But he doubts a direct link will ever be definitively made. Under normal circumstances, "40% of everybody will get cancer," he says. "It doesn't seem to me that it's possible to do an epidemiological study that will see an increased risk." Still, it may be valuable to conduct studies to reassure the population that they are not being misled, he adds.

A far greater health risk may come from the psychological stress created by the earthquake, tsunami and nuclear disaster. After Chernobyl,

evacuees were more likely to experience post-traumatic stress disorder (PTSD) than the population as a whole, according to Evelyn Bromet, a psychiatric epidemiologist at the State University of New York, Stony Brook. The risk may be even greater at Fukushima. "I've never seen PTSD questionnaires like this," she says of a survey being conducted by Fukushima Medical University. People are "utterly fearful and deeply angry. There's nobody that they trust any more for information."

Overall, the reports do lend credibility to the Japanese government's actions immediately after the accident. Shunichi Yamashita, a researcher at Fukushima Medical University who is heading one local health survey, hopes that the findings will help to reduce stress among victims of the accident. But they may not be enough to rebuild trust between the government and local residents. Tatsuhiko Kodama, head of the radioisotope centre at the University of Tokyo and an outspoken critic of the government, questions the reports' value. "I think international organizations should stop making hasty reports based on very short visits to Japan that don't allow them to see what is happening locally," he says.

UNSCEAR's working committee of roughly 70 scientists still has much to do before the final report is completed. Committee members will continue to independently validate sources of data from the accident and work on models of the flow of radioisotopes from the reactors into the environment. For the workers, "individual medical follow-up is more important than the statistical follow-up," Weiss says. "People want to know whether what we say is true." ■ SEE EDITORIAL P.415

Additional reporting by Rina Nozawa.

FACILITIES

Biosafety concerns for labs in the developing world

As report reveals lax standards in Asia-Pacific, researchers debate how to enforce rules.

BY EWEN CALLAWAY

Biocontainment labs across the Asia-Pacific region all too often fail to live up to the term. An inspection of dozens of labs has found that nearly one-third of the biosafety hoods intended to protect workers from deadly pathogens did not work properly — an offence for which a Western lab could be shut down. In one facility, only a shower curtain enclosed a table on which the brains of rabid dogs were routinely dissected.

Such deficiencies are symptomatic of a biosafety crisis in many of the laboratories that diagnose and study infectious agents in developing countries, say biorisk experts who attended a meeting at London's Chatham House on 17 May, where the results of the inspection were presented. The weaknesses could have repercussions around the globe if pathogens were released. "The strength of a chain is based on its weakest link, and developing countries are the weakest link," says Teck-Mean Chua, former president of the Asia-Pacific Biosafety Association based in Singapore, which co-sponsored the anonymized laboratory inspection.

Complaints of inadequate lab protocol in developing countries may not surprise many biologists, but they are attracting attention as scientists and research agencies in the West place increasing emphasis on biosafety. Discussions at the meeting skirted around the controversies surrounding the publication of research on mammal-transmissible forms of the H5N1 influenza virus (see page 431). However, attendees did talk about how measures to protect lab workers and contain pathogens would affect research on diseases such as flu.

In most Western countries, rules on biosafety — meant to safeguard lab workers against infection — and biosecurity, which protects the general public, became much stricter after 2001, when anthrax attacks in the United States raised the spectre of bioterrorism using laboratory-prepared pathogens. But stringent biosafety and biosecurity rules are unworkable in many developing countries, where researchers often need to handle infectious agents such as anthrax and plague to protect public health, but lack the infrastructure of the West, says Nigel Lightfoot, an associate fellow at the Centre on Global Health Security at Chatham House,



F. MAY/DPA/NEWS.COM

Biosafety measures that are mandatory in the West are not always possible in developing nations.

who chaired last week's meeting. "When you don't have any electricity, the answer is not to build a very high-security laboratory," he says. "You've got to move away from the costly bells-and-whistles solutions to what is practical."

Speakers suggested solutions such as small biocontainment boxes, for example, and also pointed out that pathogens that are endemic in a particular region present a lower biosecurity risk there than in Western laboratories. "We cannot stop them from working on things they need to for the health of their countries," says Tim Trevan, executive director of the International Council for the Life Sciences, a non-governmental organization based in Arlington, Virginia, which is interested in biosafety and biosecurity in the Middle East and Africa.

Lightfoot believes that "you're going to have dual standards" to cover different areas. But having two sets of lab rules may not sit well with either side, says Nicoletta Previsani, who heads the biorisk-management team at the World Health Organization (WHO) in Geneva, Switzerland. Scientists in developing countries may feel that they are being left with less-than-safe labs, whereas those in richer countries could feel overburdened by regulations that others don't have to follow.

➔ **NATURE.COM**
For more on the
mutant-flu furor:
go.nature.com/mhmibi

Some meeting attendees, including Chua, called on the WHO, the Food and Agriculture

Organization of the United Nations in Rome and the World Organisation for Animal Health in Paris to take the lead in establishing global standards for lab safety and security.

But Previsani says that such organizations cannot tell their member states how to operate their labs. Lightfoot adds that the WHO lacks the money and staff to act as a regulator. In January, the agency issued a five-year plan on laboratory biorisk management, in which it emphasized that it would be better placed in an organizational role, coordinating activities between stakeholders (see go.nature.com/xd9vdj).

Lightfoot argues that networks of non-governmental organizations and biosafety bodies ought to press developing nations to institute better lab standards. Donors could also help by paying more attention to the long-term sustainability of labs that they help to establish, adds Toby Leslie, an epidemiologist at the London School of Hygiene and Tropical Medicine, who has trained lab workers in Afghanistan. Too often, he says, money is spent on infrastructure and equipment without considering whether the laboratories can be operated safely by knowledgeable staff for years to come. A planned national health laboratory in Kabul, for instance, will need long-term support for maintenance and training. "I can't see a way that Afghanistan is going to be able to support it independently," says Leslie. ■

MEDICINE

US government sets out Alzheimer's plan

Health-agency budget windfall includes funding for two promising clinical trials.

BY MEREDITH WADMAN

Of the top ten leading causes of death in the United States, Alzheimer's disease — which ranks sixth — is particularly devastating in that there is no cure, no way to prevent it and no proven way to slow its progression. And with at least 11 million Americans expected to have the disease by the middle of the century (see 'Degeneration generation') — boosting the annual costs of health care to more than US\$1 trillion — the US government is anxiously looking to researchers to improve the prognosis.

Last week, the government set out how it planned to spend a \$50-million top-up to this year's funding for the National Institutes of Health (NIH) in Bethesda, Maryland, announced in February as part of a bid to "prevent and effectively treat Alzheimer's disease by 2025".

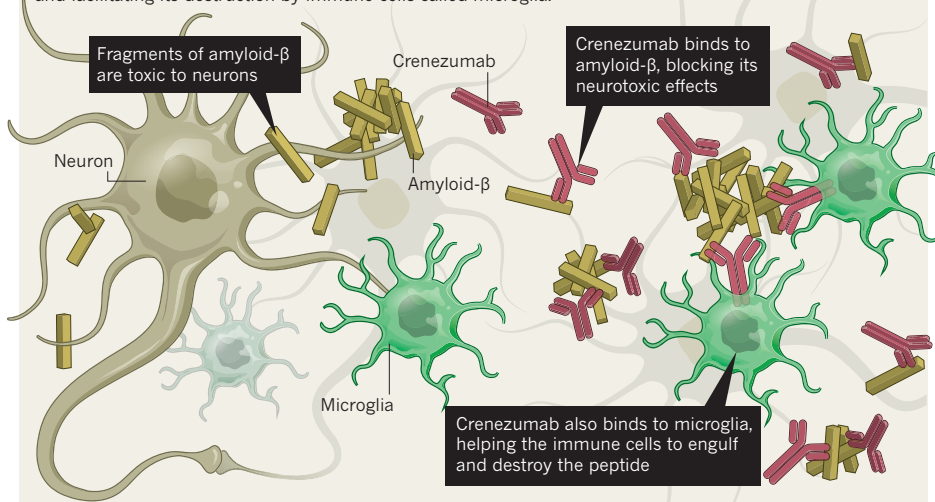
The money adds to the \$448 million that the NIH was allocated to spend on the disease this year, and roughly half of it is already being used by scientists funded by the National Human Genome Research Institute and the National Institute on Aging. They are preparing to conduct whole-genome and whole-exome studies to discover mutations that may predispose someone to the disease or protect against it. The scientists are assembling a bank of thousands of DNA samples from patients and other people whose DNA could be informative — such as elderly individuals who carry predisposing mutations but show no sign of the disease. The first results from the effort are expected as early as the end of this year.

Of the remaining money, \$4 million will fund Alzheimer's-oriented grants to small businesses, and \$2 million will be used to help generate model human Alzheimer's cells from induced pluripotent stem cells. The rest will be devoted to two groundbreaking clinical trials, one of which builds on the fact that Alzheimer's begins to attack the brain 10–15 years before symptoms appear — a conclusion confirmed by new imaging tools that allow the disease to be monitored in living people.

This trial will receive \$16 million, and is the first attempt to avert the onset of Alzheimer's by treating people who have

NEURAL DEFENDERS

The drug crenezumab, an antibody, is thought to thwart Alzheimer's disease by binding to the peptide amyloid- β , blocking its toxic interactions with neurons and facilitating its destruction by immune cells called microglia.



not yet developed symptoms. It is a five-year study using an antibody called crenezumab, made by Genentech of South San Francisco, California. Already in phase II trials in patients with mild-to-moderate symptoms, the drug is thought to work by binding to fragments of amyloid- β — insoluble, neuron-damaging peptides that aggregate in the brains of those with Alzheimer's — both neutralizing it and making it easier for immune cells in the brain

to eliminate it (see 'Neural defenders'). The Banner Alzheimer's Institute (BAI) in Phoenix, Arizona, together with a team at the University of Antioquia in Medellín, Colombia, will test the drug against a placebo in 300 symptom-free people, two-thirds of whom carry a rare genetic mutation that makes them certain to get the disease. Mild cognitive impairment sets in at a median age of 44 in people carrying the mutation, and full-blown dementia at 49, decades earlier than is common with the more typical sporadic form of the disease.

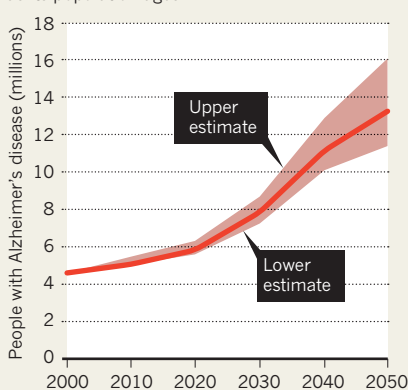
Most of the participants will come from a well-studied extended Colombian family living in Medellín and its surrounding rural areas that is estimated to have 5,000 members. They range from educated urbanites to humble farmers, all descendants of the same eighteenth-century founder. Although most do not know, and do not wish to know, their genetic status, about 1,500 are estimated to carry the mutation.

"You cannot meet these families and not be converted to their cause. It's heartbreaking," says Eric Reiman, one of the trial's three leaders and executive director of the BAI.

The other trial will receive \$7.9 million and is being led by Suzanne Craft, a neuropsychologist at the University of Washington and associate director of geriatric research at the VA Puget Sound Health Care System,

DEGENERATION GENERATION

The prevalence of Alzheimer's disease is expected to rise sharply in the United States as its population ages.



both in Seattle. It will test an intranasal insulin spray against placebo in people with mild cognitive impairment or early Alzheimer's. Insulin receptors are abundant in synapses in brain regions that are important for memory and cognition, and in Alzheimer's, amyloid- β can knock out the receptors and disrupt memory formation. Flooding rat brain cells with insulin has been found to block the fragments and protect the receptors (F. G. De Felice *et al. Proc. Natl Acad. Sci. USA* **106**, 1971–1976; 2009). When Craft tried the strategy in humans by giving intranasal insulin to people with mild cognitive impairment or early Alzheimer's, the results were impressive: about three-quarters remained stable or improved cognitively over the four months of treatment, Craft says.

PLAN FOR THE FUTURE

Craft says that she was “thrilled” to be funded for a larger, follow-on trial. But like many in the field, she worries that the national Alzheimer's plan is under-funded in the long term. “We all would like to believe that this is an important moment,” she says, “but we need to see follow-up in order to be truly convinced that this will provide an ongoing, coherent effort to address Alzheimer's disease.”

On 15 May, an advisory council to the US Department of Health and Human Services recommended that the government quickly ramp up research spending to \$2 billion per year if it wants to meet its goal of developing effective treatments and preventive measures by 2025. That would mean a quadrupling of the NIH's current spending on Alzheimer's research.

“We really have all the pieces we need to move forward in the development of effective therapeutics.”

“We really have all the pieces we need to move forward in the development of effective therapeutics,” says Paul Aisen, a neuroscientist at the University of California, San Diego, and director of the Alzheimer's Disease Cooperative Study — a 21-year-old programme of government-funded clinical studies aimed at developing Alzheimer's treatments. “The missing piece right now is the money.”

The first test of a follow-up will come later this year as Congress wrestles over the administration's 2013 budget request for the NIH. Included in the request is an additional \$80 million for Alzheimer's research. Proponents say that a failure to adequately fund research today will only lead to higher costs later.

“We're in trouble now,” says Aisen, “but in another 20 years, if things don't change, we're going to be in a dreadful situation.” ■

POLICY

Global council aims to coordinate science

Research-agency heads from around the world agree to formulate shared principles to aid collaborations.

BY RICHARD VAN NOORDEN

International research collaborations are multiplying fast, with one-quarter of the world's science and engineering publications now featuring authors from more than one country. But not all national funding agencies manage their science in the same way — researchers in China win grant funding through very different processes from their European peers, for example — which can hamper projects that span borders.

To tackle the problem, a voluntary forum, the Global Research Council (GRC), has been formed to share best practice and encourage common principles. Last week, the leaders of about 50 national research-funding agencies met at the headquarters of the US National Science Foundation (NSF) in Arlington, Virginia, to discuss the GRC's agenda: issues such as peer review, data sharing, research integrity, open access, career development and ethical conduct in research on humans. As the largest-ever gathering of research agencies, it was a “historic moment,” says Suzanne Fortier, president of Canada's Natural Sciences and Engineering Research Council.

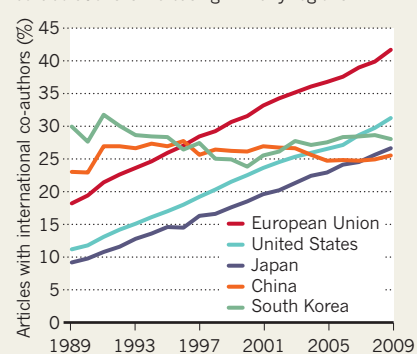
NSF director Subra Suresh, who coordinated the meeting, hopes that the GRC will broker international collaborations and co-funding arrangements to boost the globalization of science (see ‘Global science’). From now on, says Glauco Oliva, president of Brazil's National Council for Scientific and Technological Development, agencies will get together much more frequently — with at least one meeting in each of five world regions before each annual global meeting of the GRC — and will send more staff members on exchange visits abroad.

Organizing the forum was certainly an impressive feat, but it isn't yet clear how the GRC might develop. “The key thing is that this shouldn't become a big, expensive, international science talking shop,” says Kieron Flanagan, who studies science and technology policy at the University of Manchester, UK.

The GRC's first meeting produced a set of short, uncontroversial statements on common principles for peer review to assess the merit of proposed scientific projects. The principles include transparency, integrity, impartiality and confidentiality, but are not legally binding,

GLOBAL SCIENCE

The proportion of research papers from international collaborations is increasing in many regions.



SOURCE: NSF SCIENCE AND ENGINEERING INDICATORS 2012

serving only as common aspirations. “If the statements stay at this general level they're not going to be Earth-shattering,” says Flanagan.

But behind the scenes, agency heads were sharing tips and experience. Oliva, for example, says that his agency is coming under government pressure to focus on practical problems, and is rethinking its evaluation systems to include indicators that value innovation. “It's nice to see how other countries are dealing with these indicators,” he says.

Before the next major GRC meeting — in Berlin in 2013, by which time the council is expected to have almost 100 members — the group will discuss research integrity and open access to scientific data and published research. The basic principles of research integrity have already been laid down in international guidelines such as the Singapore Statement, agreed by researchers and funders in 2010. But Matthias Kleiner, who heads the DFG, Germany's main research-funding agency, says that the GRC could discuss practical questions such as how to tackle the problem of researchers being sanctioned for misconduct in one country, but continuing their research freely elsewhere.

Agreeing on principles for open access will be much tougher, says Kleiner. But it is such an important issue, he adds, that “only in a global collaboration could we come to really reliable, practical, sustainable solutions”. ■

➔ **NATURE.COM**
Read more at
Nature's metrics
special:
www.nature.com/metrics

both in Seattle. It will test an intranasal insulin spray against placebo in people with mild cognitive impairment or early Alzheimer's. Insulin receptors are abundant in synapses in brain regions that are important for memory and cognition, and in Alzheimer's, amyloid- β can knock out the receptors and disrupt memory formation. Flooding rat brain cells with insulin has been found to block the fragments and protect the receptors (F. G. De Felice *et al. Proc. Natl Acad. Sci. USA* **106**, 1971–1976; 2009). When Craft tried the strategy in humans by giving intranasal insulin to people with mild cognitive impairment or early Alzheimer's, the results were impressive: about three-quarters remained stable or improved cognitively over the four months of treatment, Craft says.

PLAN FOR THE FUTURE

Craft says that she was “thrilled” to be funded for a larger, follow-on trial. But like many in the field, she worries that the national Alzheimer's plan is under-funded in the long term. “We all would like to believe that this is an important moment,” she says, “but we need to see follow-up in order to be truly convinced that this will provide an ongoing, coherent effort to address Alzheimer's disease.”

On 15 May, an advisory council to the US Department of Health and Human Services recommended that the government quickly ramp up research spending to \$2 billion per year if it wants to meet its goal of developing effective treatments and preventive measures by 2025. That would mean a quadrupling of the NIH's current spending on Alzheimer's research.

“We really have all the pieces we need to move forward in the development of effective therapeutics.”

“We really have all the pieces we need to move forward in the development of effective therapeutics,” says Paul Aisen, a neuroscientist at the University of California, San Diego, and director of the Alzheimer's Disease Cooperative Study — a 21-year-old programme of government-funded clinical studies aimed at developing Alzheimer's treatments. “The missing piece right now is the money.”

The first test of a follow-up will come later this year as Congress wrestles over the administration's 2013 budget request for the NIH. Included in the request is an additional \$80 million for Alzheimer's research. Proponents say that a failure to adequately fund research today will only lead to higher costs later.

“We're in trouble now,” says Aisen, “but in another 20 years, if things don't change, we're going to be in a dreadful situation.” ■

POLICY

Global council aims to coordinate science

Research-agency heads from around the world agree to formulate shared principles to aid collaborations.

BY RICHARD VAN NOORDEN

International research collaborations are multiplying fast, with one-quarter of the world's science and engineering publications now featuring authors from more than one country. But not all national funding agencies manage their science in the same way — researchers in China win grant funding through very different processes from their European peers, for example — which can hamper projects that span borders.

To tackle the problem, a voluntary forum, the Global Research Council (GRC), has been formed to share best practice and encourage common principles. Last week, the leaders of about 50 national research-funding agencies met at the headquarters of the US National Science Foundation (NSF) in Arlington, Virginia, to discuss the GRC's agenda: issues such as peer review, data sharing, research integrity, open access, career development and ethical conduct in research on humans. As the largest-ever gathering of research agencies, it was a “historic moment,” says Suzanne Fortier, president of Canada's Natural Sciences and Engineering Research Council.

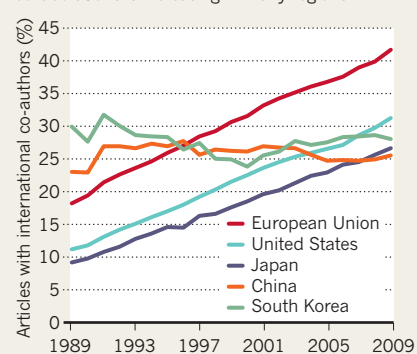
NSF director Subra Suresh, who coordinated the meeting, hopes that the GRC will broker international collaborations and co-funding arrangements to boost the globalization of science (see ‘Global science’). From now on, says Glauco Oliva, president of Brazil's National Council for Scientific and Technological Development, agencies will get together much more frequently — with at least one meeting in each of five world regions before each annual global meeting of the GRC — and will send more staff members on exchange visits abroad.

Organizing the forum was certainly an impressive feat, but it isn't yet clear how the GRC might develop. “The key thing is that this shouldn't become a big, expensive, international science talking shop,” says Kieron Flanagan, who studies science and technology policy at the University of Manchester, UK.

The GRC's first meeting produced a set of short, uncontroversial statements on common principles for peer review to assess the merit of proposed scientific projects. The principles include transparency, integrity, impartiality and confidentiality, but are not legally binding,

GLOBAL SCIENCE

The proportion of research papers from international collaborations is increasing in many regions.



SOURCE: NSF SCIENCE AND ENGINEERING INDICATORS 2012

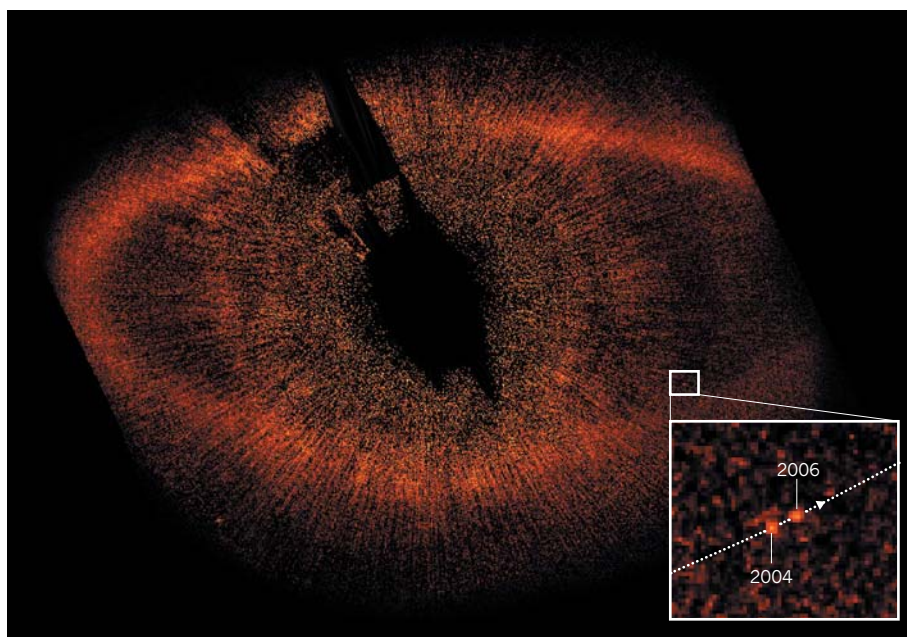
serving only as common aspirations. “If the statements stay at this general level they're not going to be Earth-shattering,” says Flanagan.

But behind the scenes, agency heads were sharing tips and experience. Oliva, for example, says that his agency is coming under government pressure to focus on practical problems, and is rethinking its evaluation systems to include indicators that value innovation. “It's nice to see how other countries are dealing with these indicators,” he says.

Before the next major GRC meeting — in Berlin in 2013, by which time the council is expected to have almost 100 members — the group will discuss research integrity and open access to scientific data and published research. The basic principles of research integrity have already been laid down in international guidelines such as the Singapore Statement, agreed by researchers and funders in 2010. But Matthias Kleiner, who heads the DFG, Germany's main research-funding agency, says that the GRC could discuss practical questions such as how to tackle the problem of researchers being sanctioned for misconduct in one country, but continuing their research freely elsewhere.

Agreeing on principles for open access will be much tougher, says Kleiner. But it is such an important issue, he adds, that “only in a global collaboration could we come to really reliable, practical, sustainable solutions”. ■

➔ **NATURE.COM**
Read more at
Nature's metrics
special:
www.nature.com/metrics



Astronomers using the Hubble Space Telescope to study the dust-shrouded star Fomalhaut (main image) hope to determine whether a bright point seen in 2004 and 2006 is a planet orbiting the star (arrow, inset).

ASTRONOMY

Hubble to revisit exoplanet puzzle

Team aims to settle running dispute over mysterious object.

BY RON COWEN

Twinkle, twinkle, little spot: are you a real world or not? Pinning down the elusive nature of Fomalhaut b, a faint speck of light discerned within the glare of a nearby star, has proved to be far from child's play. At stake are bragging rights to the first optical image of an exoplanet, and one that is close enough to be studied in detail. But since Fomalhaut b was first reported in 2008, its identity has been the subject of intense debate. The next round of Hubble Space Telescope observations, scheduled to begin on 27 May, could help to resolve the controversy.

Located a mere 7.7 parsecs (25 light years) away, the star Fomalhaut has captivated planet hunters since the mid-1980s, when infrared observations revealed that it is surrounded by dust — a possible signal of a recently formed solar system. In 2005, after further study showed that the dust took the form of a narrow band with a sharp inner edge, astronomer Paul Kalas of the University of California, Berkeley, and his colleagues suggested that a planet's gravity was shaping the dust's structure¹.

Proof seemed to come in 2008, when Kalas

and his collaborators reported that Hubble observations of the Fomalhaut star system nearly 2 years apart revealed a spot that shifted position, as would be expected of a body orbiting just inside the dust band's inner edge². From its brightness, they proposed that the object, Fomalhaut b, is a planet roughly half the mass of Jupiter. That would make it the first exoplanet to be directly imaged in visible light rather than through indirect effects, such as a wobble of the parent star. NASA was quick to flag the milestone with a public announcement.

But because of the star's extreme brilliance, says Kalas, "everything you see, even the dust belt, is extremely difficult to image". And puzzling details in the data left the matter far from settled. By the time of the second Hubble observation, in 2006, Fomalhaut b had faded to half the brightness observed in 2004 at an optical wavelength of 0.6 micrometres. Even more troubling, a preliminary search for the object in the infrared — the part of the spectrum in which newborn planets are expected to glow the brightest — failed to reveal anything.

Efforts to learn more were thwarted in January 2007, when the camera channel used by Hubble to acquire the image stopped working.

Instead, in 2010, Kalas made a third set of observations using the older Space Telescope Imaging Spectrograph on Hubble. Those results suggested that the object is heading into the dust band — not the path expected for a planet sculpting the band's inner edge. Earlier this year, another team reported that NASA's infrared Spitzer Space Telescope had failed to detect the body at all³. And on 27 April, a modelling study showed that a star might form a narrow band of dust without needing a planet to shape it⁴.

"There could still be a planet there, but the light detected has to come from something else, most likely from scattering of starlight by a large cloud of dust," says astronomer David Lafrenière of the University of Montreal in Quebec, Canada, who is a co-author of the Spitzer study.

Lafrenière notes that another recent study supports that idea. Images from the European Space Agency's Herschel Space Observatory reveal that the band around Fomalhaut is packed with fine dust, which the star's radiation pressure continually sweeps into interstellar space. For the dust band to be maintained, the study authors conclude, the equivalent of 2,000 one-kilometre-wide comets must be destroyed in the ring every day. Such destruction could create a compact cloud of dust that would shine like the object seen by Hubble, says Lafrenière.

Adding to the intrigue, Christian Marois and Raphaël Galicher of the Herzberg Institute for Astrophysics near Victoria in British Columbia, Canada, have reanalysed the original Hubble data and say they cannot confirm that Fomalhaut b has faded or that it is veering into the dust band. Marois adds coyly that a much bigger surprise could be lurking in the analysis, but he won't unveil details until mid-June, when Kalas plans to release the results of the next Hubble observations.

Those spectrograph images should help to pin down the orbit of Fomalhaut b and whether it varies in brightness, Kalas says. But to determine whether the object is truly a planet, he adds, Hubble will have to monitor it for several more years to see whether it behaves as a planet, a dust cloud or something else.

Some researchers, however, have already consigned Fomalhaut b to the dustbin. "Some people are bending over backward to keep the planet alive. There's no justification any more," says Ray Jayawardhana, an astronomer at the University of Toronto in Ontario, Canada. Along with Lafrenière and other collaborators⁵, he is co-discoverer of another object, J1609, which he says is the first true exoplanet to be imaged directly.

Over to you, Hubble. ■

1. Kalas, P., Graham, J. R. & Clampin, M. *Nature* **435**, 1067–1070 (2005).
2. Kalas, P. *et al. Science* **322**, 1345–1348 (2008).
3. Janson, M. *et al. Astrophys. J.* **747**, 116 (2012).
4. Lyra, W. & Kuchner, M. J. Preprint at <http://arxiv.org/abs/1204.6322> (2012).
5. Lafrenière, D. *et al. Astrophys. J.* **719**, 497 (2010).

CLIMATE CHANGE

Cancelled project spurs debate over geoengineering patents

SPICE research consortium decides not to field-test its technology to reflect the Sun's rays.

BY DANIEL CRESSEY

Technologies to keep Earth cool could one day provide a radical fix for climate change — and, in a world struggling to control its greenhouse-gas emissions, could also prove highly lucrative for inventors.

But should individual researchers, or companies, be allowed to own the intellectual property (IP) behind these world-changing techniques? The issue was thrust into the spotlight last week after a controversial geoengineering field trial was cancelled amid concerns about a patent application by some of those involved in the project, as first reported by *Nature*¹.

The £1.6-million (US\$2.5-million) Stratospheric Particle Injection for Climate Engineering (SPICE) project was funded by the UK government to investigate whether spurting reflective aerosols into the stratosphere could help to bounce some of the Sun's warming rays back into space. As part of this project, SPICE had planned to test a possible delivery system: pumping water up a 1-kilometre-long hose to a balloon, where it would be sprayed into the sky.

The project had already sparked protests from environmentalists wary of geoengineering². But “a potentially significant conflict of interest” over a patent application for SPICE's technology, which some team members only recently became aware of, was a decisive factor in the cancellation, says project leader Matthew Watson, an Earth scientist at the University of Bristol, UK. The patent was submitted by Peter Davidson, a consultant based on the Isle of Man who was an adviser at the workshop that gave rise to SPICE, and Hugh Hunt, an engineer at the University of Cambridge, UK, who is one of the SPICE project investigators.

UK funding bodies require anyone assessing or applying for grants to declare relevant

potential conflicts of interest. Davidson and Hunt say that they were clear about their patent application before SPICE was awarded funding, and there is no suggestion that they acted inappropriately. But at least one of the funding councils is now investigating the circumstances surrounding the SPICE grant, and the patent in question, says Watson.

Hunt blames a culture clash for the confusion. “It is completely normal for engineering projects to be protected by IP,” he says. “The issue here is that in climate science there is mistrust of IP, and I understand that now.” He says he does not expect to earn any money from the patent.

SPICE's climate modelling and other technology development work will continue, but the incident is another blow for a field already troubled by concerns over governance. In 2010, researchers and policy-makers gathered at the Asilomar Conference Center near Monterey, California, to agree a set of guiding principles for the field — an effort that largely failed³.

The following year, a smaller group produced the ‘Oxford Principles’, stating that geoengineering should be “regulated as a public good”. The lead authors of those principles later warned that patenting of geoengineering technologies could “have serious negative impacts”, by creating a culture of secrecy that could delay much-needed developments.

Climate scientist David Keith of Harvard University in Cambridge, Massachusetts, agrees, advocating legal restrictions on patents related to solar-radiation management. Any technologies that could be controlled by a small number of people, yet have the capacity to rapidly alter our planet's climate, “are deeply troubling”, he says. But Keith is not against

patenting in principle — he has applied for patents on techniques to remove carbon dioxide directly from the atmosphere.

Shobita Parthasarathy, a public-policy researcher at the University of Michigan, Ann Arbor, says that the field urgently needs to agree on detailed rules for IP. In 2010, she noted a “dramatically increasing” number of patent applications in the area, containing broad language that could allow a small number of patent holders to take control of a huge swathe of technologies⁴. One possible solution, she says, is to develop a unique system for handling geoengineering patents, akin to the way that atomic-energy patents are controlled in the United States. That system puts certain technologies off-limits, and allows the government to take control of some intellectual property. “I don't think the solution is to get rid of IP,” she says.

Another option might be to allow patent-holders to receive royalties, but without the option to restrict the use of the patent, says Tim Kruger, a researcher at the Geoengineering Programme, University of Oxford, UK, who helped to develop the Oxford Principles. This would allow some research and development to proceed, while still providing a financial incentive to work in the area, he says.

But geoengineering patents of any kind could give companies a vested interest in the continuation of climate change, argues Holly Buck, a social scientist who has studied the ethics of geoengineering. “It seems conceptually wrong to create conditions for an enterprise that would institutionally benefit from a stressed climate.” ■ **SEE EDITORIAL P.415**

1. Cressey, D. *Nature* <http://doi.org/hw2> (2012).
2. Macnaghten P. & Owen R. *Nature* **479**, 293 (2011).
3. Tollefson, J. *Nature* **464**, 656 (2010).
4. Parthasarathy, S. et al. *A Public Good? Geoengineering and Intellectual Property* (Univ. Michigan, 2012); available at go.nature.com/scrl4g

“The issue here is that in climate science there is mistrust of IP.”



**MORE
ONLINE**

TOP STORY

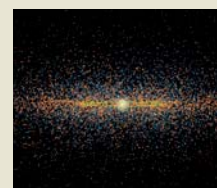


Finches learn even when practice isn't perfect
go.nature.com/vqnol

MORE NEWS

- Source found for missing water in sea-level rise go.nature.com/damx2m
- Why great ideas tend to come when you aren't trying go.nature.com/ekb8rh
- Volcanic eruptions trigger shocking finding go.nature.com/falz3m

ON THE BLOG



Tally of most hazardous asteroids doubles
go.nature.com/inyngq

THE BIOSECURITY OVERSIGHT

THE FIGHT OVER MUTANT FLU HAS THROWN THE SPOTLIGHT ON A LITTLE-KNOWN GOVERNMENT BODY THAT OVERSEES DUAL-USE RESEARCH. SOME ARE ASKING IF IT WAS UP TO THE TASK.

BY BRENDAN MAHER

The packages that started arriving by FedEx on 12 October last year came with strict instructions: protect the information within and destroy it after review. Inside were two manuscripts showing how the deadly H5N1 avian influenza virus could be made to transmit between mammals. The recipients of these packages — eight members of the US National Science Advisory Board for Biosecurity (NSABB) — faced the unenviable task of deciding whether the research was safe to publish.

The group deliberated. Soon, the rest of the NSABB's 22 voting members and two dozen non-voting members and advisers were drawn in. For five-and-a-half weeks, they pored over

the data in the papers, weighing the benefits of sharing the information against the risk that doing so might lead to the accidental or intentional release of a lethal new virus. They exchanged views in hundreds of e-mails and in more than 24 hours of teleconference calls.

On 21 November, the NSABB recommended that journals should redact the papers, publishing their conclusions but sharing methods and data only with approved scientists and health officials. It was the first time that the board had recommended any such restriction since it was convened in 2005, and it sparked a global debate — aired in journals, meetings, blogs and newspapers — that is still raging and has left the US government in an awkward spot. “The

United States funded this research and then wanted to censor it,” says David Fidler, who teaches international law at Indiana University Bloomington. “This looked dysfunctional.”

Throughout these turbulent months, the spotlight has shone as much on the NSABB as it has on the mutant flu viruses. The board's members, with backgrounds ranging from biology to medicine to national security and law, have been developing guidelines for biosecurity oversight for nearly seven years. The flu research was a major test of the principles they had been espousing.

By all appearances, the board struggled. By mid-February, the NSABB was under pressure to overturn its initial assessment. And

in the last days of March, it did — voting unanimously in favour of full publication for one paper, which appeared early this month¹. The board also recommended that the second paper be published, but six members dissented, arguing that the work still posed significant concerns. (That paper's publication is expected within weeks.) The whole episode has left many people with questions. Could the board have done better? Why wasn't the research flagged earlier? And is there a way to publish sensitive information while minimizing risks?

There is one point of agreement, says David Relman, a microbiologist at Stanford University in California and member of the NSABB: "This is not the way any of us wants to see these issues discussed, that is, at the eleventh hour and fifty-ninth minute."

SECURITY SCARE

The NSABB's roots can be traced back to October 2001, when letters carrying anthrax spores were sent to several public figures around the country (see "Threat and response"). In response, the US government invested billions of dollars to prepare for future acts of bioterror, much of it channelled into pathogen research through the National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda, Maryland. In parallel, Congress asked the National Academies to form a panel to recommend how dual-use research — work that could carry bioterror risks as well as benefits — should be identified, regulated and reported. Scientists were anxious to show that they could police their own work and avoid heavy-handed or cumbersome regulation from above. "The science community ought to come up with a process before the public demands the government do it for them," warned Parney Albright of the US Department of Homeland Security in 2003 (ref. 2).

Geneticist Gerald Fink at the Massachusetts Institute of Technology in Cambridge was chosen to chair the panel. The recommendations in the resulting 'Fink report', published in 2004, set out seven 'deadly sins': types of research that should warrant close scrutiny, such as experiments to render a vaccine ineffective or to make a pathogen more virulent. The report also called for the creation of a national advisory board to further explore the issues on a national and international stage. This would become the NSABB, an independent panel that is managed and supported by the National Institutes of Health (NIH). In June 2005, NIH director Elias Zerhouni swore in 23 NSABB members in Bethesda. Paul Keim, a microbiologist at Northern Arizona University in Flagstaff and acting chair of the NSABB, says that the ceremony involved the raising of hands. "We all kept from giggling," he says.

Right away, the board started to flesh out guidelines for a US policy on dual-use research. Its flagship document, released in 2007 and building on the Fink report, emphasized

local self-governance, suggesting, for example, that investigators monitor their own and colleagues' projects, possibly with the help of existing institutional biosafety committees.

Although not officially part of the board's remit, the NIH also called on the NSABB to review the occasional paper that raised biosecurity concerns. The first two^{3,4} to land in the board's lap, in 2005, dealt with efforts to resurrect the Spanish flu virus that was responsible for millions of deaths immediately after the First World War. The board recommended that the papers be published in full. Keim says he now wishes that the group had had more time to deliberate over the Spanish flu work, which raised many of the same issues as the

"THE UNITED STATES FUNDED THIS RESEARCH AND THEN WANTED TO CENSOR IT."

current debate. "I guess I have some regrets about that decision because of the impact it would have had on policy," he says.

Nevertheless, the papers the board received last October were different from those it had handled before. Their roots go back to 1997, when H5N1 started devastating bird populations worldwide and health officials voiced alarm about the catastrophe that could ensue if the disease gained the ability to jump between humans. In 2006, the NIH convened a blue-ribbon panel to identify priority research on avian influenza. Among other projects, it highlighted the need for experiments to see how bird flu might evolve the ability to spread from person to person. Soon after, the NIH commissioned and funded several such projects, including one from Ron Fouchier at the Erasmus Medical Center in Rotterdam, the Netherlands, and one from Yoshihiro Kawaoka at the University of Wisconsin-Madison and the University of Tokyo. Robert Webster, a virologist at St Jude Children's Research Hospital in Memphis, Tennessee, and a member of the blue-ribbon panel, says that it paid close attention to the stringent biosafety requirements of such work, but that dual-use concerns "didn't really surface".

They should have, says Keim. The experiments committed at least two of the Fink report's deadly sins: they deliberately changed the host range of a pathogen and they increased its transmissibility. "You think about adapting H5N1 to mammals," Keim says, and you quickly "realize that there is the potential to do something very dangerous".

Concerns surfaced in September 2011, when Fouchier presented his results at a high-profile meeting in Malta. He described, in ominous terms, how he had mutated wild H5N1 virus to make it more likely to infect human cells. He had then let the virus evolve in ferrets, a good model for human transmission, until it was able to spread through the air by a cough or a sneeze. Kawaoka took a different approach, mutating a single gene from H5N1 and plugging it into a less pathogenic viral genome. What resulted — two influenza viruses that could spread in mammals, that most humans had never been exposed to and that stemmed from a virus with the potential to kill — was worrying.

Still, the board struggled with its decision. At first, says NSABB member Arturo Casadevall, a microbiologist at the Albert Einstein College of Medicine in New York, "I was very uncomfortable with the idea of redacting information because I think that it's a slippery slope". But the data and expert analysis assembled by the board convinced him that what Fouchier and Kawaoka had done was too easy to repeat. "We just didn't think it would be a good idea to put a recipe out there," he says. Michael Osterholm, a public-health researcher at the University of Minnesota in Minneapolis, emphasized his support for the research, but stressed the precautionary principle. Once the work was published, it could not be taken back. "You can't unring a bell," he said on several occasions.

In late December, the US Department of Health and Human Services, which oversees the NIH, announced that it would follow the NSABB's advice. The response was severe, says Keim. "That redaction approach has been universally panned," he says. "The investigators hated it, the people who weren't going to get the data hated it. The government hated it because they couldn't figure out how to do it."

Meanwhile, the NSABB's members were scrambling to make clear that the issues needed international discussion. In mid-February, Kawaoka and Fouchier presented their work at a closed meeting at the World Health Organization (WHO) in Geneva, Switzerland. They assured the researchers that the benefits — for monitoring wild viruses for potentially dangerous mutations and for vaccine development — outweighed the risks. They also explained that the mutant viruses weren't necessarily lethal to the ferrets, something that hadn't been clear to everyone before. The attendees, mostly academic flu researchers, recommended that both papers be published in full.

In light of the new information, the NIH asked the NSABB to reconsider its position. A workshop was scheduled for 29–30 March.

SECOND THOUGHTS

The meeting started at 7 a.m. in a sixth-floor conference room of building 31 on the NIH campus in Bethesda. Keim had heard the presentations at Geneva, but still couldn't predict how the rest of the board were going to react.

THREAT AND RESPONSE

Long-standing concerns about research with potential risks erupted into full-blown controversy late last year. The National Science Advisory Board for Biosecurity (NSABB) has been a central player throughout.

2001 Australian researchers inadvertently create a highly pathogenic mousepox virus, prompting alarm that the technique could be used to weaponize smallpox.



2001 Anthrax attacks in the United States show the reality of a bioterror act using sophisticated microbiology.

2001

2002



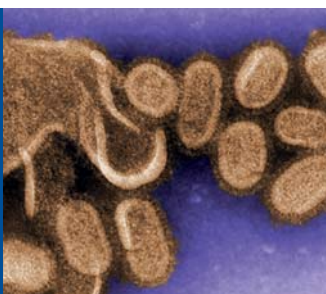
2003 A highly pathogenic form of H5N1 avian influenza begins to circulate. Several people are infected and as many as half the confirmed cases die from it.

2003 The National Academies publishes the 'Fink' report (*Biotechnology Research in an Age of Terrorism*) which calls for the creation of the NSABB.

2002

2004

2005 The NSABB reviews two papers reconstructing the 1918 Spanish influenza virus genome. It recommends full publication of both.



2005 The NSABB is sworn in.

2006 A blue-ribbon panel publishes a report prioritizing research into H5N1, which leads to funding of controversial work by Ron Fouchier and Yoshihiro Kawaoka.

2007 The NSABB publishes guidelines entitled *Proposed Framework for the Oversight of Dual Use Life Sciences Research*.

2005

2006

2007



12 September Ron Fouchier (pictured) announces that he has created a mammalian-transmissible form of H5N1.

21 November The NSABB recommends that papers by Fouchier and Kawaoka be redacted: only certain researchers would gain access to full materials and methods.

Controversy on mutant avian flu begins

20 January Fouchier, Kawaoka and 37 other flu researchers agree to a 60-day moratorium on research with the mutant strains.

17 February Experts convened by the World Health Organization conclude that the research should be published in full.

30 March After a two-day meeting, the NSABB recommends full publication of both papers, but the decision is not unanimous.

2 May Kawaoka's paper is published.

2011

2012

P. SCHNEIDER, THE CHARLOTTE OBSERVER/AP; GETTY IMAGES; T. TUMPEY/CDC; D.-J. VISSER/THE NEW YORK TIMES/REDUX/EYEVINE

"I was not placing bets either way." The voting members sat round a conference table, with about 60 administrators, government officials and ex officio members looking on. Everyone was given two hours in silence to review revised manuscripts from Kawaoka's and Fouchier's teams. The researchers had edited the papers to clarify the benefits of the research and to explain the safety measures taken during work with the viruses. Later, they gave presentations. Fouchier was reportedly questioned for two hours.

By this point it was clear that Kawaoka's paper posed less of a threat than Fouchier's because of the low pathogenicity of his hybrid virus. But Relman and other members of the NSABB say that they were not reassured by Fouchier or by the revisions to his manuscript. "There were no new data that for me diminished the evidence for mammal-to-mammal transmissibility and no data that convinced me that the virulence was any less in his mutant viruses than it was in the wild-type parental H5N1 strains," Relman says.

The board also heard that the practical and political barriers to redaction looked formidable. NIH director Francis Collins told them that export-control rules and freedom-of-information laws in other countries would make it impossible to implement a system for selectively releasing data quickly. Moreover, such a system could jeopardize the pandemic-influenza preparedness framework, an international agreement to share influenza viral samples and information that had been hammered out in 2011 by the WHO after years of debate. For officials in countries such as Indonesia, where poultry farmers have faced financial ruin because of H5N1, a decision to redact information sounded like a decision to withhold it. It became clear to the board that redaction was effectively off the table, meaning that the NSABB could vote to publish the paper in full, or not at all.

After a full day of briefings and another of deliberation, the board voted. The members present at the meeting unanimously recommended publication of the Kawaoka paper and voted 12–6 in favour of publishing Fouchier's.

Few came out of the meeting happy. Some were still unsure about how dangerous Fouchier's virus really was. "Even the 12 who voted in favour of publication were uneasy about this uncertainty in the virus," says Keim, who declined to reveal his vote. Relman, who voted against publication, says that the process felt unbalanced and that he didn't have enough time to assess some new data presented there that had not yet been peer reviewed. "I do think questions should be asked about the manner and process by which we were asked to perform this reassessment," he says.

Osterholm asked some of these questions in a

NATURE.COM
Read more about
the mutant-flu
controversy:
go.nature.com/mhmibi

sharply worded letter to the NSABB and Amy Patterson, the board's director at the NIH, a week and a half after the meeting. (The letter was leaked to *Science* and *Nature* days later.) In it, Osterholm said that the presentations given at the meeting were one-sided and designed to favour full publication of the articles. He said that Fouchier had revealed at the meeting an additional mutation that makes H5N1 both transmissible through the air and deadly. This work "surely must be considered as a candidate for the next manuscript to be before the NSABB for review," wrote Osterholm, who worried that all the same problems would come up again. In her response to the letter, Patterson respectfully disagreed with

**"THIS IS NOT THE
WAY ANY OF US
WANTS TO SEE
THESE ISSUES
DISCUSSED."**

Osterholm's complaints. But by this point, the spat had started to attract the attention of law-makers. Congressman Jim Sensenbrenner (Republican, Wisconsin) wrote letters to the NIH and to the White House asking how decisions about the research were reached.

People within the NSABB, and outside it, now say that the board did its best in a highly complex situation. But many point a finger at a flawed mechanism for identifying and dealing with dual-use research. "Almost at every step the system isn't working very well for these projects that raise serious concerns about biosecurity," Fidler says. The most pressing question is why the research wasn't flagged up earlier for scrutiny.

The answer: the policy simply wasn't in place. In its 2007 report, the NSABB recommended that the federal government develop guidelines and implement a code of conduct to help institutions and researchers to report potential risks at the earliest stage of project development. It also recommended the development of strategies for communicating sensitive research, including restricted publication. These recommendations went largely unheeded because scientists resisted the introduction of cumbersome new practices. "We got all worried about the possibility of these threats," says Fidler, but when it came to imposing regulations on research, "we tended to back off".

Now, the flu controversy has forced the US government's hand. On 29 March, while the NSABB was being briefed, it released a policy

that requires federal agencies to identify and monitor research projects they fund that tick boxes on the 'deadly sins' list. Tom Inglesby, who directs the Center for Biosecurity of UPMC in Baltimore, Maryland, welcomes the new policy. "It would be much more preferable for these decisions to go on at the beginning of this experimental process. It's more fair to the scientists, more fair to their institutions, more fair to the journals and more fair to the NSABB," he says.

Keim, however, points out that the policy does not require review by disinterested parties. "These are decisions that need to be made in the open with input from different segments of our society," he says. It may be too much to expect scientists to coolly evaluate the risks of their own research against the benefits they gain personally from publication. And even if regulatory changes do take root in the United States, international agreement will take years to solidify. Keim and several others at the NSABB say that publishing with controlled access to certain data would still have been the preferred option for the H5N1 papers, but the challenges extend well past US borders.

Most observers and participants expect that the NSABB will continue to weigh in on policy development, although it may have to resolve questions about conflicts of interest first. In the wake of the flu controversy, some observers have questioned whether it is appropriate to have the NSABB under the control of the NIH — which funded the flu research — and populated by NIH-funded scientists. Board members might not have wanted to vote against publication if it risked biting the hand that feeds them. "I'd be lying if I didn't say that that thought crossed my mind," says Michael Imperiale, a virologist at the University of Michigan in Ann Arbor and a member of the board since its inception. Ultimately, he says, he followed his conscience, which favoured publication of both articles. Anthony Fauci, director of the NIAID and a non-voting member of the NSABB, calls the idea of the NIAID taking revenge against NSABB members for their vote "preposterous".

The whole controversy has been an ordeal for those involved. But Casadevall takes a positive view. "The end result has been a tremendous education," he says.

"I don't know how much of a silver lining that is," Fidler says. There's little consensus as to what a new system for dual-use research oversight should look like, he says, and governments have simply kicked the can down the road in the past. "That may happen again, but at least it's out in the open," he says. ■

Brendan Maher is a features editor for *Nature*.

1. Imai, M. *et al. Nature* <http://dx.doi.org/10.1038/nature10831> (2012).
2. Check, E. *Nature* **421**, 197 (2003).
3. Tumpey, T. M. *et al. Science* **310**, 77–80 (2005).
4. Taubenberger, J. K. *et al. Nature* **437**, 889–893 (2005).



CHAMBER OF PHYSICS

The world's largest underground laboratory has been a success story for Italian science. But 30 years after construction began, its future is uncertain.

The drive along Italy's highway A24 from the central Adriatic coast towards Rome begins with a winding climb into the snow-covered Apennine Mountains, followed by a plunge into the 10-kilometre-long tunnel under Gran Sasso, the highest peak in the region. About half way through the tunnel, a detour leads off to the right. It reaches a dead end almost immediately at a heavy iron gate. But press the intercom button and utter the words 'particle physicist' into the microphone, and the gate slides open like something from a James Bond movie.

Not far beyond the gate is a car park. From there one continues on foot, and begins to get some idea of the scale of the infrastructure hidden beneath the mountain. Opening off a long corridor are three huge halls, each about 20 metres wide, 18 metres high and 100 metres long. This vast area is the home of the Gran Sasso National Laboratory, part of the Italian National Institute of Nuclear Physics (INFN).

In fact, the laboratory's 180,000 cubic metres of space is not its most valuable attribute. Lying under 1,400 metres of rock, it offers silence — not an absence of sound, but of cosmic-ray noise, the rain of particles

BY NICOLA NOSENGO

constantly bombarding Earth's surface from space. This lack of cosmic interference has attracted a generation of physicists to these halls, where they can study some of the rarest and most elusive phenomena in the Universe.

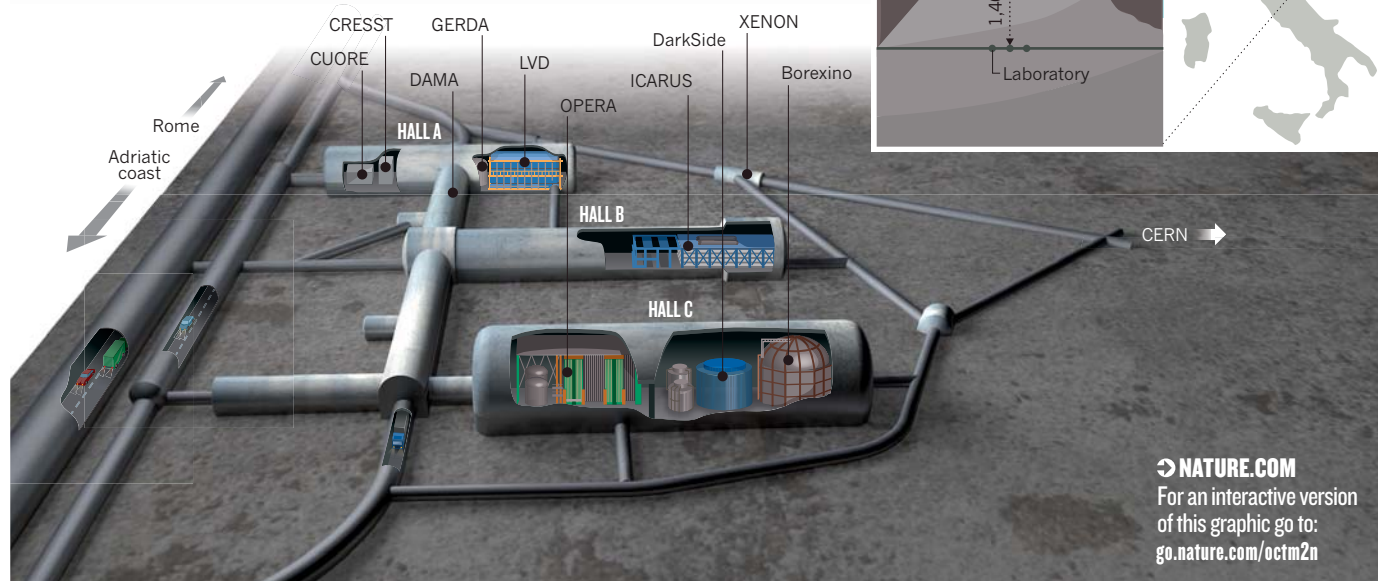
Most people probably first heard of Gran Sasso last September, when its OPERA experiment reported — incorrectly, as it turned out — that neutrinos seemed to travel faster than light. But the laboratory, construction of which began 30 years ago, has long been known to physicists. Gran Sasso "was the first true underground laboratory, the only one purposely built for science", says Stanley Wojcicki, a physicist at Stanford University in Palo Alto, California. It is still by far the largest, serving as a base for 18 experiments and about 950 researchers from 32 countries.

"Gran Sasso's halls have allowed experiments based on different technologies to work side by side, comparing each other's pros and cons, and building multiple generations of the same experiments," says Kevin Lesko, a neutrino physicist at the Lawrence Berkeley National Laboratory in Berkeley, California. The result has been some

Gran Sasso's Large Volume Detector (LVD) searches for neutrino bursts from supernovae.

THE A, B AND C OF GRAN SASSO

Experiments at the Gran Sasso National Laboratory are housed in and around three huge halls carved deep inside the mountain, where they are shielded from cosmic rays by 1,400 metres of rock.



long-standing underground rivalries. But they have helped to make Gran Sasso one of Italy's strongest scientific success stories, responsible for a string of notable results in neutrino and solar physics. "It is our trading currency with the international physics community" says INFN president Fernando Ferroni.

At the age of 30, however, Gran Sasso finds itself in transition. Its scientific priorities are changing, and a long collaboration with CERN, Europe's premier particle-physics laboratory near Geneva in Switzerland, is nearing its end. Budget cuts are making it harder to keep Gran Sasso running. International competition is increasing, thanks to proposals to put in new detectors at other facilities such as SNOLAB near Sudbury in Canada, Japan's Kamioka laboratory and the Soudan laboratory in Minnesota, as well as the proposed Deep Underground Science and Engineering Laboratory at the Homestake gold mine in South Dakota. And although many of its experiments have had important results, Gran Sasso is still waiting for its first big discovery: a groundbreaking finding that would match its size and the ambitions that spawned its creation.

THE MYSTERY OF THE MISSING NEUTRINOS

When Gran Sasso was first conceived in the late 1970s, one of the key motivations was political. Antonino Zichichi, who proposed and oversaw the project as president of the INFN, admits he wanted to give Italy something that would increase its weight in the international physics community and counterbalance CERN's role in European physics.

But there were also two solid scientific drivers. The first was the case of the 'missing' neutrinos. Since the late 1960s, US physicist Raymond Davis had been running an experiment in Homestake to detect neutrinos created in nuclear reactions at the core of the Sun. He found only about one-third of the number predicted by theory. Other physicists had suggested a solution: if neutrinos 'oscillate' as they travel in space, spontaneously transforming from one of their three types to another, they would be so thoroughly mixed by the time they arrived at Earth that only about one-third would still be the type that Davis's detector could pick up. But because neutrinos interact so weakly with other particles, testing that idea would require a detector shielded from cosmic rays by hundreds of metres of rock.

The other problem was proton decay. Various efforts to build 'grand unified theories' of particle interactions suggested that the proton was ever so slightly unstable, and would decay into lighter subatomic

particles with a half-life many times longer than the age of the Universe. The only hope of seeing such a rare event was to monitor a huge amount of matter for a long time, while shielding it from any background radiation that might swamp the signal.

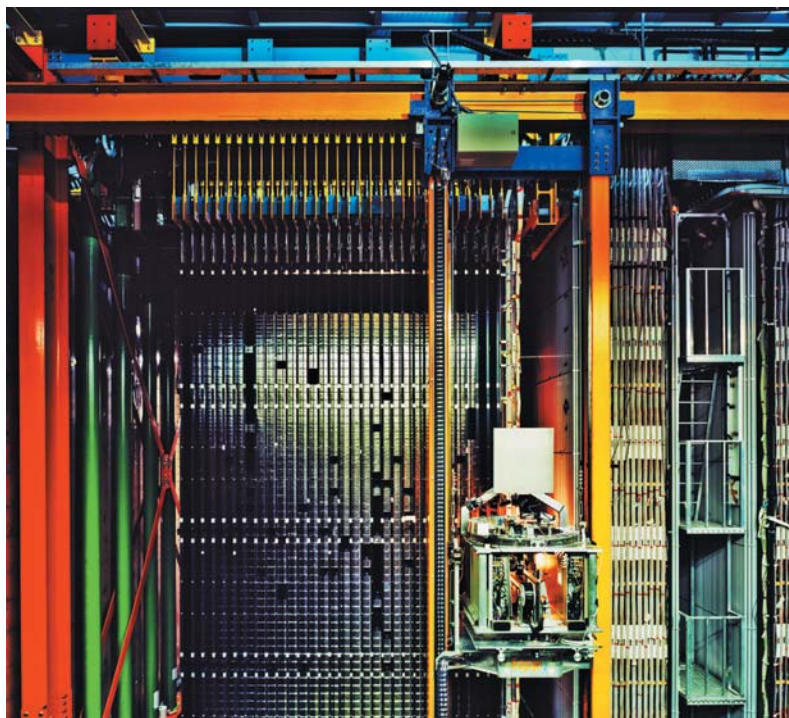
All the project needed was a gigantic cavern in a mountain. Zichichi found it at Gran Sasso, 120 kilometres from Rome, where a highway tunnel under the mountain had been started and then abandoned for economic and political reasons. "The shape of the mountain was ideal", he recalls. "Gran Sasso is flat, so the thickness of the shield remains constant along its length." INFN physicists measured cosmic rays in the gallery, and found that their flux was about one million times lower than at the surface. Radioactivity measurements were also encouraging, one thousand times lower than on the ground. Gran Sasso's dolomite rock is very poor in uranium and thorium, the main sources of natural radioactivity.

A mountain with the ideal physical properties, already containing a tunnel but devoid of traffic — "it seemed too good to be true", says Zichichi. Had they had to build the laboratory from scratch, he says, it would have cost too much. As it was, its estimated price tag came to 77 billion Italian lire, roughly equal to US\$160 million today.

Zichichi presented the project to the Italian Parliament in 1979. It was approved in February 1982 — a decision had also been made to complete the abandoned highway — and construction began in September. In 1987, the halls were ready for the installation of the first experiments.

Building a scientific community took longer. "For a long time, Gran Sasso was mostly a training camp for underground physics," says Ferroni — an entirely different endeavour from particle accelerator physics, which was where most of the physicists coming to work in Gran Sasso had learned their craft. Rather than smashing particles against each other and seeing what happens, underground physicists spend their days painstakingly selecting and testing materials for use as shields and detectors, weeding out every conceivable source of background noise, and analysing months and months of data — only to find that, nine times out of ten, a 'signal' is just background noise that somehow got through.

The first-generation of experiments established what could and could not be done with underground detectors. Neutrino physics proved especially fruitful: one of Gran Sasso's most notable early experiments was Gallex, which ran from 1991 to 1997 and provided confirmation that Davis's absent solar neutrinos really were missing¹. But the effort to measure proton decay was quickly abandoned. "Early theories were



At Gran Sasso, neutrinos are under intense scrutiny, with the OPERA (left), Borexino (top right) and ICARUS experiments all focused on their detection.

too naive,” says Ferroni. “Seeing a proton decay would require one million tonnes of water, or more than 10,000 tonnes of liquid argon. We lack the technology, the laboratory, the money.” Also shot down was the quest for magnetic monopoles: theoretical particles with just one magnetic pole, which were the subject of Gran Sasso’s first major experiment, a primarily Italian–US collaboration called MACRO. Researchers searched for the particles between 1989 and 2000, but never caught a glimpse of them².

OPERA TAKES THE STAGE

By the 2000s, Gran Sasso had entered into a phase marked by new scientific priorities, a fully formed scientific community and the construction of the larger, more ambitious experiments that now fill its halls (see ‘The A, B and C of Gran Sasso’). Coming from the car park, one first encounters Hall C and the Borexino experiment, a follow-on from Gallex that studies the full energy range of neutrinos emitted by the Sun. Borexino is a huge dome, 18 metres in diameter, with two concentric spheres nested inside it like Russian dolls. The outer dome is filled with ultrapure water that helps to reduce the radiation reaching the innermost sphere, which detects flashes of light caused by neutrinos hitting a scintillator liquid.

Cristiano Galbiati, a physicist at Princeton University in New Jersey and a member of the Borexino team, says that the experiment’s unmatched background-noise reduction is what allowed the team to report last October the first ever detection of low-energy ‘pep’ neutrinos, which are produced by a rare reaction in the Sun whereby protons and electrons interact to make deuterium³. Predicted by theory but never previously observed, pep neutrinos formed one of the last missing pieces in models that explain how the Sun keeps burning. “Low-energy neutrinos from the Sun oscillate in a completely different way from high-energy ones,” says Galbiati. “The only way to solve the solar neutrino problem once and for all is to measure neutrinos across the whole spectrum, and this is the only experiment that has knocked background noise down enough.”

“GRAN SASSO IS
OUR TRADING
CURRENCY WITH
THE INTERNATIONAL
PHYSICS
COMMUNITY.”

Right beside Borexino is the now famous OPERA experiment: a detector the height of a three-storey building made of about 150,000 ‘bricks’ of photographic film separated by lead plates. Since 2008, OPERA has been looking for oscillations in a beam of neutrinos fired from CERN, some 730 kilometres away. The beam comprises only muon neutrinos, and the idea is to look for tau neutrinos at the Gran Sasso end, which would prove that they have switched from one type, or ‘flavour’, to another during their journey. OPERA reported its first tau neutrino in 2010 (ref. 4), and the team expects to confirm one or two more before the end of 2012.

But that achievement was not all that proponents had hoped. The experiment was first proposed in the 1990s, but negotiations with CERN, whose management was wary of distracting staff and funds from its conventional work, dragged on for the better part of a decade. By the time OPERA finally began collecting data, neutrino oscillations had long since been confirmed by similar neutrino-beam experiments in Japan and the United States. The one consolation was that those experiments were only able to record muon neutrinos disappearing from their beams — so far, only OPERA has detected their reappearance as taus.

However, those results were eclipsed by the OPERA team’s announcement in September that its neutrinos seemed to be travelling faster than light⁵. The resulting media frenzy exacerbated divisions among team members — some of whom refused to sign the announcement preprint, arguing that the collaboration had no business publicizing what was almost certainly an experimental error. The dissidents’ objections were also echoed early and often in neighbouring Hall B, home to OPERA’s arch-rival ICARUS, a pair of rectangular tanks each 20 metres long and filled with 300 tonnes of liquid argon, that competes with OPERA on the detection of tau neutrinos. “OPERA’s confidence level on tau neutrinos is not very high,” says deputy ICARUS spokesperson Sandro Centro. “We will report our first results this year and we think we can do better.” In October, the ICARUS team flatly contradicted OPERA with its own neutrino-velocity measurement, finding that the particles do not travel faster than light⁶.

In the end, the dissidents were right: the source of the error, a loose data cable, was disclosed in February. But the incident led to the resignations of both OPERA's spokesperson, Antonio Ereditato, and its physics coordinator, Dario Autiero. Former Gran Sasso director Mario Monacelli, who is now an OPERA scientist and one of those who did not sign the velocity paper, says that the frenzy has not interfered with the experiment's day job. "But it has created tensions that will take some time to go away," he says. "We are now working to prevent the collaboration from breaking into opposing factions."

Another locus of tension can be found in Hall A and its surroundings, much of which are occupied by experiments searching for dark matter. The existence of dark matter is abundantly clear to astronomers, who can see its gravitational effect on galaxies and clusters of galaxies through their telescopes. But its nature is a mystery: the stuff is utterly transparent, and passes through stars and planets as if they weren't there. A prevailing theory holds that it is a haze of weakly interacting massive particles (WIMPs) that formed during the Big Bang, and have permeated the Universe ever since. The trick is to catch and study WIMPs in a laboratory detector — a task that, once again, requires ultra-low background radiation.

Gran Sasso's dark matter (DAMA) experiment, which opened in 1996 and has gone through many upgrades since, looks for flashes of light that occur when dark-matter particles collide with the atoms of a sodium iodide crystal. DAMA works on the assumption that Earth's velocity through the dark matter in our Galaxy varies as it orbits the Sun, producing an annual variation in the flux of dark-matter particles passing through the detector. For 13 years, the team has been reporting just such an oscillation in signal, although one that is consistent with dark-matter particles lower in mass than most theorists expect⁷.

The source of the tension lies just a few metres away, in XENON100: a US-led experiment that has been using more than 100 kilograms of liquid xenon as a detector for WIMPs since 2009. With its first publication in 2010, covering its first 11 days of data, the XENON team found no sign of WIMPs in DAMA's mass range⁸. The public finger-pointing goes on to this day. DAMA's Rita Bernabei insists that XENON's results depend on a specific model of dark matter. "There are candidates, scenarios and uncertainties that can account for all the results presented so far," she says, including those from both DAMA and XENON. But XENON scientist Francesco Arneodo is dubious. "You can play with models as much as you want, but it's a pity they did not make any effort to stimulate similar experiments elsewhere, which would close the argument," he says.

Ferroni sees only two possible resolutions for the DAMA controversy. "Either they will end up describing some previously unknown source of background noise," he says, "or they will go straight to Stockholm" to collect a Nobel prize.

THE ROAD AHEAD

Leaving behind the close atmosphere of the underground halls and following the highway tunnel to its western end, it's a short drive to the base of the mountain. Here, in Gran Sasso's above-ground offices, director Lucia Votano is planning its future.

Funding remains a concern, she says. Gran Sasso costs the INFN almost €10 million (US\$12.8 million) a year — not including the experiments. The agency's budget has shrunk by one-third during the past decade, and a further 5% cut is anticipated in 2013 owing to Italian austerity measures.

But even without cuts, she says, many things will soon change under the mountain. By the end of 2012, the neutrino beam will vanish when CERN shuts down its accelerators to upgrade its flagship Large Hadron Collider. OPERA and ICARUS will almost certainly be disbanded. "At the moment it's really not clear what the future of neutrino physics in

Europe will be," says Centro — although the ICARUS detector will be moved to CERN and used as a testbed to develop much larger detectors.

Gran Sasso, meanwhile, will be left to focus on dark matter, and on a second longstanding mystery just as crucial for understanding phenomena not predicted by the standard model of particle physics: neutrino-less double beta decay. This is a hypothetical form of radioactivity, thought to be extremely rare, that would be possible only if the neutrino is its own antiparticle. In 2001, Hans Klapdor-Kleingrothaus and his colleagues reported detection of the phenomenon using a now-dismantled experiment in Gran Sasso⁹. But no other experiment has been able to replicate the finding so far, and Klapdor-Kleingrothaus's use of statistics has been criticized by some scientists¹⁰. Although, admits Ferroni, who counts himself a sceptic, "if those results are wrong, they are so in a much less obvious way than the OPERA one".

Votano and Ferroni expect that Gran Sasso's GERDA experiment, which is looking for neutrino-less double beta decay in a detector with exceptionally low background radiation, will decisively confirm or reject the result within a couple of years. In around

2014, it will be joined by CUORE, a tower made from 1 tonne of tellurium dioxide. In the end, both experiments will probably prepare the ground for even larger detectors, which will further increase the probability of spotting the very unlikely event they are after. "It is too early to say which technology works best," says Votano.

As for dark matter, the XENON team is now building a 1-tonne detector — expected to begin collecting data by late 2014 — that will become the world's most sensitive detector for WIMPs. Another experiment in the making, DarkSide, will use 50 kilograms of liquid argon as its detector and will borrow the successful background reduction system of the Borexino experiment.

DAMA, meanwhile, will undergo a further upgrade. And the German experiment CRESST is already looking for dark-matter collisions in crystals at temperatures near absolute zero.

Given all that, says XENON spokesperson Elena Aprile, a physicist at Columbia University in New York, "the first detection of a WIMP will come from Gran Sasso". Or, at least, she says, "this is where we will understand if we are completely on the wrong track".

However, they have competition. Similar dark-matter and double-beta-decay experiments are under construction or already running at Homestake, Sudbury and Kamioka. And with the completion of the 2,500-metre-deep Jin-Ping laboratory two years ago, China has joined the dark-matter race. Many of these experiments may take advantage of better natural shielding — something that physicists measure as metres of water equivalent. Gran Sasso's shielding equals 3,300 metres of water, against 4,100 at Homestake and almost 6,000 at Sudbury.

On the other hand, says Votano, Gran Sasso has developed techniques — such as the Russian-doll system pioneered by Borexino — that allow it to simulate a greater depth. It remains the largest and best equipped underground laboratory in the world. And, Aprile says, pointing to the snowy peaks that surround the external laboratories, the location is hard to beat. "If I can make good physics in a place like this, why not?" ■

Nicola Nosengo is a freelance writer based in Rome.

"EITHER THEY WILL
END UP DESCRIBING
SOME PREVIOUSLY
UNKNOWN SOURCE
OF BACKGROUND
NOISE OR THEY WILL
GO STRAIGHT TO
STOCKHOLM."

1. Anselmann, P. *et al. Phys. Lett. B* **285**, 376–389 (1992).
2. Ambrosio, M. *et al. Eur. Phys. J. C* **25**, 511–522 (2002).
3. Bellini, G. *et al. Phys. Rev. Lett.* **108**, 051302 (2012).
4. Agafonova, N. *et al. Phys. Lett. B* **691**, 138–145 (2010).
5. Adam, T. *et al. Preprint at* <http://arxiv.org/abs/1109.4897> (2011).
6. Antonello, M. *et al. Preprint at* <http://arxiv.org/abs/1203.3433> (2012).
7. Bernabei, R. *et al. Eur. Phys. J. C* **67**, 39–49 (2010).
8. Aprile, E. *et al. Phys. Rev. Lett.* **105**, 131302 (2010).
9. Klapdor-Kleingrothaus, H. V., Dietz, A., Harney, H. L. & Krivosheina, I. V. *Mod. Phys. Lett. A* **16**, 2409–2420 (2001).
10. Aalseth, C. E. *et al. Mod. Phys. Lett. A* **17**, 1475–1478 (2002).

COMMENT

ENVIRONMENT Removal of oestrogens from water will cost billions **p.441**

RESOURCES How land increases social and economic disparity **p.442**

PHYSICS An illustrated history of the atomic bomb **p.445**



DRUG POLICY Introduce categories for regulating probiotics **p.446**

NAT. CANCER INST./CORBIS



The golden age of antibiotic discovery began with systematic testing of soil microbes by Selman Waksman (pictured centre).

Recover the lost art of drug discovery

Bacterial evolution is overwhelming our antibiotic defences, says **Kim Lewis**. Using modern technology to replicate past success might tip the balance in our favour.

The more we know about antibiotics, the fewer we can discover. This is essentially why the field is in trouble — only one antibiotic belonging to a new class, the narrow-spectrum daptomycin, was discovered and made it into clinical practice in the past 50 years. After decades of little success, pharmaceutical firms are pulling the plug on antibiotic discovery. Most have either left the field (such as Merck, based in New Jersey, and Eli Lilly in Indiana) or are downsizing their effort. But the spread of antibiotic-resistant pathogens continues unabated¹.

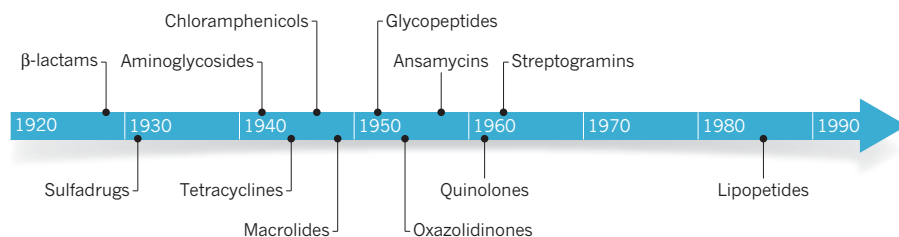
Advances in medicine have exacerbated the problem. Clinicians are struggling to treat infections that tolerate antibiotics, in which bacteria produce a small quantity of dormant, spore-like persister cells². The only function of dormant persisters is survival; once the antibiotic concentration drops, persisters wake up, start propagating and the infection relapses. Although this type of infection is rare in healthy individuals, whose immune cells can effectively kill both growing and dormant persister cells, it has become an increasing problem as a side effect of

medical intervention. Cancer chemotherapy and transplant drugs, for instance, weaken the immune system, and artificial devices — catheters, heart valves and orthopaedic implants such as hip joints — are ideal platforms on which biofilms can form. Immune cells cannot penetrate these mats of bacteria, so they become a protective habitat for dormant persisters.

The increase and spread of antibiotic-resistant pathogens — which have evolved resistance mechanisms against specific drugs — is another cause for alarm³. Given the ►

ANTIBIOTIC DISCOVERY TIMELINE

Decades without identifying antibiotics that go on to be used for the treatment of patients has put our defence against bacteria at risk. This timeline pinpoints the year that the antibiotics were first discovered.



► lack of new antibiotics, we are in a stand-off with human pathogens. And we are poised to lose, prompting a return to the pre-antibiotic era of epidemics. Several human pathogens have already produced globally resistant strains, such as *Acinetobacter baumannii* that causes non-healing wounds, and some strains of tuberculosis (*Mycobacterium tuberculosis*). Although these pathogens represent only a tiny proportion of the bacteria out there, resistance is increasing. For example, highly resistant strains of the superbug MRSA (methicillin-resistant *Staphylococcus aureus*) were initially restricted to hospitals, but are now widely present in the community.

Why did we lose the ability to discover antibiotic compounds? I think that we should look to the golden era of discovery, in the middle of the last century, and revive what worked then. Several approaches were successful in the past but were abandoned, each for its own reason. However, these can now be revitalized using modern tools and expertise. Now is the time to recover the lost art of antimicrobial drug discovery.

RETURN TO OUR ROOTS

Following the discovery of penicillin in 1928, the golden era for finding new antibiotics really began in the 1940s, when microbiologist Selman Waksman exploited bacteria's ability to produce their own antibiotics (which they use to out-compete each other) by systematically testing soil microbes — largely streptomycetes. This led to the discovery in 1943 of streptomycin, the first antibiotic used to treat tuberculosis. The method that Waksman used — an approach I will refer to as the Waksman platform — was widely adopted by the pharmaceutical industry and yielded the main classes of antibiotics over the following 20 years (See 'Antibiotic discovery timeline'). But despite years of success, this overmining of soil bacteria resulted in diminishing returns — the same compounds were continuously found, and the platform collapsed.

This period also saw the identification of the main classes of synthetic compounds, but only a single broad-spectrum class — the fluoroquinolones — was introduced in the 1970s and 1980s. The main reason why so few synthetic compounds work is because they cannot penetrate the bacterial cell envelope.

Scientists can identify antimicrobials from screening, or by designing compounds that fit into active sites of bacterial enzymes and inhibit their functions, but they cannot cross this formidable barrier.

More than a decade ago, medicinal chemist Christopher Lipinski found a way to circumvent another problem that was plaguing drug discovery: how to find compounds that could be absorbed well by the human gastrointestinal tract, making them available for oral use. With his team, Lipinski culled information from hundreds of approved drugs and came up with a list of five properties (such as size and water-repellent nature) that made compounds more prone to absorption. Since then, chemists have synthesized compounds using the Lipinski rule of five to ensure that they can be taken orally. I believe that we should do the same for antibiotics, by developing 'rules of penetration' that enable antibiotics to break through the sturdy bacterial cell envelope. The number of known molecules that penetrate well into bacteria is very small, not enough to form a set of rules, but we can measure how well compounds from a collection of randomly selected chemicals enter bacteria, rank these, and deduce the desirable properties for penetration. Knowing these rules should then enable us to build focused libraries of compounds that are likely to reach their targets in bacterial cells.

OLD IDEAS

We should also revive the approach that allowed the discovery of several compounds in the 1950s that were effective against *M. tuberculosis*. At that time, tuberculosis was still a major cause of death in the Western world, and to identify treatments, scientists screened synthetic compounds against *M. tuberculosis*, discovering isoniazid, pyrazinamide, ethionamide and ethambutol. Now, researchers screen libraries to identify compounds that work against many different pathogens, and those compounds are much harder to find. Therefore, unsurprisingly, since the four synthetic compounds specific to *M. tuberculosis* were identified no others have emerged despite researchers screening a thousand times more molecules when the global library increased to more than ten million compounds. Although physicians

will always need broad-spectrum antibiotics, they could also use targeted compounds, particularly since the rapid diagnosis of the pathogen causing an infection is becoming available. This would allow selective therapy, using a pathogen-specific compound that will leave the body's natural protective flora untouched.

In addition, scientists should attempt to synthesize 'prodrugs' again. First discovered in the 1950s, these compounds are harmless until they enter a bacterial cell, at which point an enzyme converts them into a toxic substance, with properties similar to bleach. Prodrugs can kill both growing and dormant cells, suggesting that they could be effective against the seemingly invincible persister cells that make infections so difficult to treat. In the 1960s, reasoning that compounds without specific targets might be toxic, scientists began testing all compounds to see if they hit specific targets in bacterial cells. However, this resulted in the discarding of prodrugs, which do not have specific targets, just general toxicity to bacteria once converted. This explains, at least in part, why the discovery of synthetic compounds, many of which are prodrugs, ceased in the 1960s. I think that we should look again at prodrugs, because they may be a powerful weapon against both relapsing and untreatable chronic bacterial diseases.

Finally, we need to revive the Waksman platform. We stopped discovering useful classes of antibiotics from soil bacteria decades ago, but that was during a time when it was impossible to culture most (more than 99% of) bacterial species. Today, methods to culture these microbes are being developed, providing access to chemical diversity that was previously missed⁴. Examining the gene-expression patterns of the bacteria exposed to newly isolated antibiotics and comparing these with the characteristic expression profile of known antibiotics might avoid 'rediscovering' penicillin or streptomycin. A new pattern of gene expression will suggest a new compound, and could even point to its target.

Opportunities for antibiotic discovery are certainly there. What is not known is how effectively we will take advantage of them. ■

Kim Lewis is director of the antimicrobial discovery center, Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA. K.L. sits on the board of directors at NovoBiotic Pharmaceuticals in Cambridge, Massachusetts, and at Arietis Corporation in Boston. e-mail k.lewis@neu.edu

1. Bush, K. et al. *Nature Rev. Microbiol.* **9**, 894–896 (2011).
2. Alekshun, M. N. & Levy, S. B. *Cell* **128**, 1037–1050 (2007).
3. Lewis, K. *Annu. Rev. Microbiol.* **64**, 357–372 (2010).
4. Lewis, K., Epstein, S., D'Onofrio, A. & Ling, L. L. *J. Antibiot. (Tokyo)* **63**, 468–476 (2010).

The hidden costs of flexible fertility

Urgent public debate is needed over a European proposal to regulate environmental levels of the active ingredient in birth-control pills, say **Richard Owen** and **Susan Jobling**.

More than 100 million women worldwide use contraceptive pills¹. The active ingredient in most formulations is ethinyl estradiol (EE2), which is excreted from the body along with other, naturally occurring, oestrogens, passing through waste-water works into rivers, estuaries and lakes.

Decades of research have shown that EE2 and other oestrogens cause widespread damage in the aquatic environment by disrupting endocrine systems in wildlife. This includes a condition called intersex: the irreversible development of eggs in the testes of male fish, which reduces their reproductive success¹. When researchers introduced EE2 into a Canadian lake in 2001 at the vanishingly low level of 5 parts per trillion, the population of one fish species collapsed². The potency of EE2 as an endocrine disrupter makes it a serious threat to wildlife and fisheries.

Governments across the globe have been painfully slow in their response to this issue. In 2004, some 30 years after the first observations of intersex in fish in British rivers, the UK Environment Agency agreed that there was a strong case for risk management³. But it was only in January this year that the European Commission announced its intention to regulate EE2 under the Water Framework Directive (see go.nature.com/8fm3dz). European countries would be required by 2021 to limit EE2 in water bodies to an annual average of no more than 0.035 parts per trillion.

This regulation will set a global precedent for regulating pharmaceuticals in the environment. It also presents society with a difficult dilemma, of which the public remains mostly unaware.

THE CONTROL DILEMMA

The pharmaceutical and water industries, and many governments, are strongly opposed to this proposed regulation. One reason is the high cost of compliance. The maximum permissible concentration of EE2 would be very low, reflecting the level at which damage occurs to aquatic organisms. The only currently effective option for removing EE2 from waste water to allow compliance involves adsorption on to



Treating waste water to remove oestrogens will cost European countries billions of euros.

granular activated carbon. For a UK town of around 250,000 people, such a system would cost more than €8 million (US\$10.3 million) to install and around €800,000 a year to operate¹. For the 1,400 waste-water works that would need upgrading in England and Wales alone, this would amount to more than €30 billion in total.

These costs — for EE2 and possibly other pharmaceuticals in the future — will be borne by the public through higher water prices. Are we willing to pay, or would we rather settle for environmental harm as collateral damage associated with flexible fertility? The answer to this dilemma is not obvious, either in Europe or worldwide.

Difficult decisions that affect us all should include us all. But there has been little open debate as to how environmental pharmaceuticals such as EE2 should be regulated. A workshop attended by representatives of governments, industry and environmental groups was convened by the European Parliament on 24 April to discuss EE2 and other 'priority substances' that are slated for regulation. But by not including the public, the European Commission is falling short of its own recommendation that "all interested parties should be involved

to the fullest extent possible" in assessing risk-management options⁴. The UK Royal Commission on Environmental Pollution has also stressed that such decisions must take into account peoples' values, and that "it is no longer acceptable for decisions to be negotiated privately between the regulator and polluter"⁵.

The need to protect our environment from the harmful effects of EE2 is clear, but understanding of our willingness as a society to pay for that protection is not. Nor is it obvious where responsibilities lie, including whether pharmaceutical companies have a moral duty of care for all their products, which could be better designed so that they are safe for the environment.

On 6 November 2012, a European Parliament legislative committee will vote on whether EE2 should be included in the priority substances list for regulation, and whether the legislation should progress to a first reading in the European Parliament next January. These decisions must be democratic, equitable and legitimate. The public must be informed about the scientific evidence, and the costs of action or inaction openly debated. A decision on whether or not to regulate EE2 must include the people it will affect, whatever the outcome. ■

Richard Owen is chair in responsible innovation at the University of Exeter Business School, Exeter EX4 4PU, UK.

Susan Jobling is chair in environmental toxicology and head of the Institute for the Environment at Brunel University, Uxbridge, Middlesex UB8 3PH, UK. e-mail: r.j.owen@exeter.ac.uk

1. Jobling, S. & Owen, R. Ethinyl oestradiol: bitter pill for the precautionary principle in: *Late Lessons from Early Warnings: Science, Precaution, Innovation* (European Environment Agency, in the press).
2. Kidd, K. A. et al. *Proc. Natl Acad. Sci. USA* **104**, 8897–8901 (2007).
3. Gross-Sorokin, M. Y., Roast, S. D. & Brighty, G. C. *Environ. Health Perspect.* **114** (S1), 147–151 (2006).
4. European Commission. Communication from the Commission on the Precautionary Principle COM (2000)1 (2000).
5. Royal Commission on Environmental Pollution Report no. 21: *Setting Environmental Standards* (RCEP, 1998).

R. BROOK/SPL



J. MORGAN/PANOS

An Indonesian child surveys ancestral lands logged to make way for an oil-palm plantation — just one example of 'land-grabbing' by outside concerns.

DEVELOPMENT

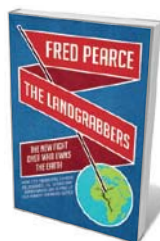
Striking out for new territory

Wendy Wolford hails a perceptive take on the people and organizations behind modern big land acquisitions.

Leo Tolstoy asked an eerily prescient question in his 1886 short story *How Much Land Does a Man Need?* That question and its answer — enough to bury a coffin, said Tolstoy — are captured perfectly in the compelling and well-researched *The Landgrabbers*. In it, science journalist Fred Pearce dissects the modern rush to acquire land for production, investment, speculation or preservation, focusing on the people and organizations with the will, and the money, to do it.

The amount of land being bought and sold worldwide has increased by a factor of at least ten in the past five to ten years, according to the World Bank. The international community became acutely aware of the scale of acquisitions in 2009, after the South Korean firm Daewoo Logistics entered a deal to buy one-third of the arable land in Madagascar — one of the world's poorest countries — to grow palm oil and maize (corn). There was widespread protest leading to a change in Malagasy government and the

deal fell through — but hundreds more continue. Media and scholarly attention to the issue is increasing. Yet understanding of these deals is still sketchy. Pearce points out that it is nearly impossible to quantify land acquisition accurately. What we do know is that land deals have complex drivers, including increasingly volatile food and fuel prices; expectations that agricultural and energy commodities will soon become scarce, given that known land and oil reserves are likely to run out within decades; rising demand for grain and meat; enhanced opportunities and incentives for financial speculation; and concerns about



The Landgrabbers: The New Fight Over Who Owns the Earth
FRED PEARCE
Beacon/Eden
Project Books:
2012. 336/400 pp.
\$27.90/£20

global environmental degradation and species loss. The hunt is on for access to land.

To find out who is behind the land deals, Pearce travelled the world for a year, investigating transactions and interviewing buyers, sellers, local community members, government officials, law students and conservation scientists. He draws on his own experience as well as a range of research (including papers from an April 2011 conference in Brighton, UK, which I co-organized as part of the Land Deals Politics Initiative).

Pearce's focus on buyers and their financial or corporate lineages is what sets *The Landgrabbers* apart. They make up an extraordinary group: asset strippers, Saudi Arabian sheiks, a "self-styled wild man of Wall Street", Christian evangelists, venture capitalists, White Russian nobles, agricultural ministers and clothing tycoons turned conservationists. Pearce documents the lofty ambitions and near-missionary zeal that drive buyers to snap up land everywhere from the semi-arid, sparsely populated Paraguayan Chaco to potential crop fields in war-torn South Sudan.

What emerges is paradoxical. On the one hand, as Pearce shows, land deals are backed by a wide variety of intentions, some of them good. There are deals intended to help Africans to develop a modern, profitable agricultural sector; to grow rice or oil palm; to build game parks from Kenya to Cambodia; to preserve land in Patagonia; and to provide South African farmers with land in Mozambique, Zambia and Georgia.

➔ **NATURE.COM**
To read about the links between oil and poverty, see:
go.nature.com/nk7emp

On the other hand, most of the deals tend to leave local people worse off, without rights to the land they have worked for generations and without access to shelter, jobs or natural resources. Pearce illustrates this point with many examples, including Dominion Farms, which he says is draining productive wetlands in western Kenya to grow rice, using chemicals that pollute the Yala River and providing only a fraction of the promised jobs — while prohibiting locals from hunting and fishing. Although the World Bank insists that land deals can be win-win for buyers and locals, most land grabbers are motivated by the bottom line. Their underlying goals are to provide benefits for an external community, through commodity production, mineral or fossil-fuel extraction or environmental protection. As Pearce points out, buyers tend to make ambitious promises that rarely come true, and land grabs seem to exacerbate the economic and social inequalities that make them possible in the first place.

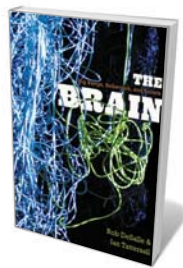
Pearce leaves some questions unanswered. He does not explain how 'land grabs' should be defined: does any purchase constitute grabbing? And how different are today's acquisitions from colonization and the 'scramble for Africa' more than a century ago? Is it possible that the international community has learned enough to stop the worst abuses of land grabbing?

The World Bank has proposed a set of voluntary principles that would counsel land grabbers to be more humane, but these have met with scepticism. Olivier De Schutter, the UN special rapporteur on the right to food, has likened such principles to pretending that "destruction of the global peasantry" can be done responsibly. But there has also been a sustained effort by social movements and the Food and Agriculture Organization of the United Nations to publicize an alternative, human-rights-based approach to land access, beginning not with how to feed and fuel the world, but rather with how to improve the well-being of people living in poverty.

Those efforts, says Pearce, could promote a future reliant on sustainable, efficient production from smallholdings. There is ample evidence that a growing world population would be better fed by improving the yields of those already on the land than by introducing large-scale, energy-intensive industrial farming. And then all we would need to 'grab' would be a cup of shade-grown coffee — and a little room for the coffin that Tolstoy mentioned. ■

Wendy Wolford is a professor of development sociology at Cornell University in Ithaca, New York, and a founding member of the Land Deals Politics Initiative at the International Institute of Social Studies of Erasmus University Rotterdam in the Netherlands.
e-mail: www43@cornell.edu

Books in brief



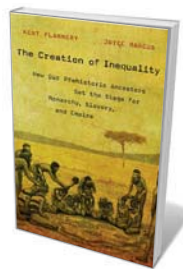
The Brain: Big Bangs, Behaviors, and Beliefs

Rob DeSalle and Ian Tattersall YALE UNIV. PRESS 320 pp. £29.95 (2012)
Dream-weaver, computer, "evolutionary mess": however we view the human brain, thinking about thinking never palls. Here, two eminent curators from New York's American Museum of Natural History — anthropologist Ian Tattersall and genomicist Rob DeSalle — cherry-pick the latest research in evolutionary biology, molecular biology and neuroscience to fashion a considered take on why humans have reached cognitive supremacy. The resulting brainfest covers the Big Bang, the evolution of nervous systems, the senses, data processing, emotions, memory, language, behaviour and more.



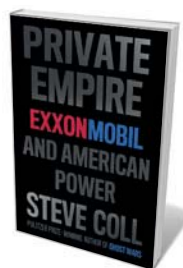
The Spark of Life: Electricity in the Human Body

Frances Ashcroft ALLEN LANE 352 pp. £25 (2012)
Physiologist Frances Ashcroft celebrates the "body electric" — animal electricity — in this interweaving of research breakthroughs, science history and human stories. Membrane proteins found in all cells, known as ion channels, produce electrical signals that control processes including the human heartbeat, vision and sexual attraction through electrical 'events' in neurons and muscle cells. Ion channels also have a role in insulin secretion, and among her fascinating offerings, Ashcroft recounts how her own findings in this area have improved treatment of a rare genetic form of diabetes.



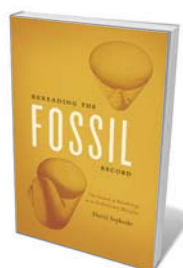
The Creation of Inequality: How Our Prehistoric Ancestors Set the Stage for Monarchy, Slavery, and Empire

Kent Flannery and Joyce Marcus HARVARD UNIV. PRESS 544 pp. £29.95 (2012)
The extreme disparities in prosperity and status that plague societies and spawn empires had their roots 4,500 years ago, argue anthropological archaeologists Kent Flannery and Joyce Marcus. Drawing on ethnography and archaeology, they reveal how societies create elites and tyrants. High rank for the ambitious and talented, for instance, becomes problematic when translated into hereditary rights, and competition among the privileged can lead to despotism.



Private Empire: ExxonMobil and American Power

Steve Coll PENGUIN 704 pp. \$36 (2012)
Oil giants don't come much bigger than ExxonMobil or, claims Pulitzer-prizewinning journalist Steve Coll, more secrecy-shrouded. On the basis of more than 400 interviews, Coll explores the years from the 1989 Exxon Valdez oil spill to 2011 — a period when the business's net cash flow was, he says, US\$493 billion — and tracks ExxonMobil's accrual of power, from Equatorial Guinea and Iraq to its Texas headquarters. A vast cast, including world leaders, corporate scientists and influential former chief executive Lee Raymond, enriches the deeply researched proceedings.



Rereading the Fossil Record: The Growth of Paleobiology as an Evolutionary Discipline

David Sepkoski UNIV. CHICAGO PRESS 440 pp. \$55 (2012)
In the 1970s, a new kid on the block was shaking up palaeontology, geology and biology. Historian David Sepkoski charts the rise of palaeobiology from 1945 to 1985, driven by a small but illustrious band of palaeontologists including Stephen Jay Gould and David Raup, who grappled with how the geological record could produce evidence for evolution. The solution, as Sepkoski engagingly relates, lay in quantitative analysis of evolutionary patterns in fossils.

BEHAVIOURAL ECOLOGY

Design for living

Manfred Milinski hails the long-awaited update of a classic on the optimal design of behaviour.

Natural selection produces impressive physical adaptations, but it is equally powerful in fashioning behaviour, given that it favours the individuals best able to avoid predators, choose mates, find food, and more. Ecology is the stage on which selection of the fittest occurs, and the discipline of behavioural ecology deals with the optimal design of behaviour.

The long-awaited update to a classic in this field is now here, presenting new directions in thinking and addressing burning questions. Richly informed by progress in many other disciplines, such as sensory physiology, genetics and evolutionary theory, it marks the emergence of behavioural ecology as a fully fledged discipline.

In 1981, John Krebs and Nick Davies's *An Introduction to Behavioural Ecology* (Blackwell Scientific Publications) boosted a flourishing new area of thought. It became fashionable to use economic theory to understand animal behaviour — for instance, the trade-off between the need to forage efficiently and to avoid being eaten.

That first edition was a masterpiece of scientific writing, embraced by behavioural ecologists worldwide. After two further successful editions at six-yearly intervals, we now have the fourth, co-authored by Stuart West and carefully updated throughout.

Among the elements appearing for the first time is a fascinating chapter on cognition, which looks at the ability of birds to plan for the future, and to behave as if they are able to interpret the knowledge of individuals that they observe. Another section shows that in evolutionary 'arms races' between, say, hosts and parasites, each invention by the parasite selects for a counter-invention by the host, which would otherwise lose the race. The result is much like the Red Queen's race in Lewis Carroll's *Through the Looking Glass*, where the pace must be kept up simply to stay in the same place. This is beautifully demonstrated using the coevolution of cuckoos and their hosts.

Studying animal personalities is trendy, but why is it advantageous for one individual to have personality A and another to

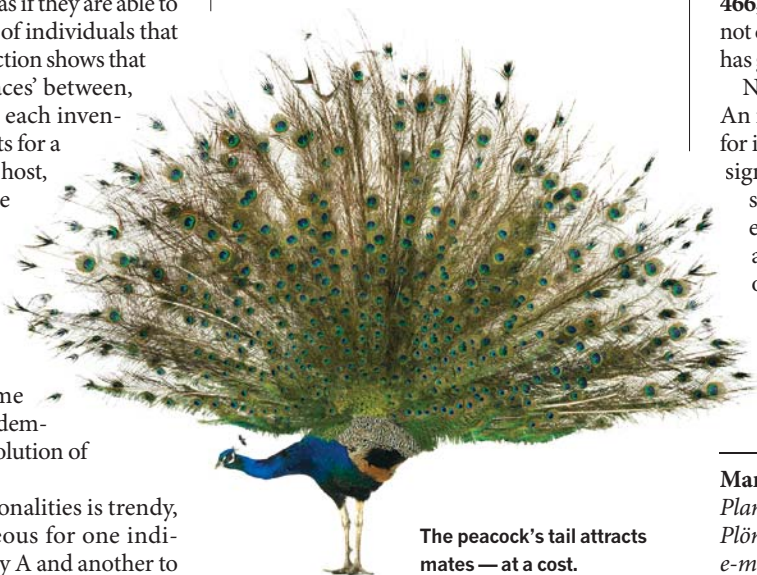
have B? Related questions addressed in the book are whether animal groups such as packs or herds have leaders, or whether it is better for the group members to 'vote' on a course of action.

The current bandwagon in behavioural ecology is sexual selection — which Charles Darwin wrote about but did not solve. This concept has produced a plethora of findings and ideas, which Davies *et al.* describe and discuss. They note that the 'runaway process', proposed by evolutionary biologist R. A. Fisher in the early twentieth century to explain why females such as peahens prefer ornamented and thus handicapped males, still awaits rigorous testing. However, the 'good genes for resistance' explanation, put forward by Bill Hamilton and Marlene Zuk in the 1980s, seems now to be confirmed. The Hamilton–Zuk mechanism posits that sexual ornaments such as the peacock's tail signal genetic resistance against the prevalent disease affecting that species, on the grounds that ill individuals could not produce such ornaments. Choosy females therefore obtain this resistance for their offspring.



An Introduction to Behavioural Ecology (Fourth Edition)

NICHOLAS B. DAVIES,
JOHN R. KREBS &
STUART A. WEST
Wiley-Blackwell: 2012.
520 pp. £34.99,
\$149.95



The peacock's tail attracts mates — at a cost.

In vertebrates, as the book describes, these 'good genes' are now known to be alleles of the polymorphic genes of the major histocompatibility complex (MHC), which direct the immune response to new infections. MHC genotypes can be detected through odour and are chosen before ornament. So a female will choose males whose odour indicates that their MHC alleles provide the optimal complement to her own. Among these, say the authors, she will prefer the most ornamented males, whose alleles include those for disease resistance. But this explanation still leaves many questions unanswered. For example, how does a Hamilton–Zuk mechanism work in insects, which lack MHC genes?

Sexual conflict also drives sperm competition, the evolution of behaviour that maximize the chances of sperm from a particular male fertilizing the egg. Will males that produce more sperm win out, or do eggs 'choose' spermatozoa with good genes? Today's science fiction is tomorrow's textbook wisdom.

The book also delves into the evolution of cooperation and altruism. Why invest resources to increase another individual's fitness? One explanation, kin selection, is that the individual helps only close relatives, which share copies of its genes. The conditions for altruistic behaviour by kin selection are given by Hamilton's rule: the benefit-to-cost ratio must be greater than one divided by the relatedness of the altruist to the recipient. So for full siblings, which have a relatedness of one-half, the benefit must be at least twice the cost. As discussed in the book, this rule seems to explain numerous phenomena, such as the social insects with their sterile worker castes, and also why we prefer to raise our own children.

There is overwhelming evidence for kin selection, but not all researchers agree with the theory behind Hamilton's rule, as a recent critique shows (M. A. Nowak *et al.* *Nature* **466**, 1057–1062; 2010). That Davies *et al.* do not discuss this dramatic development, which has great potential for new insight, is a pity.

Nevertheless, there is much here to inspire. An intriguing chapter on communication, for instance, discusses the major problem of signalling between animals — what keeps signals reliable and honest? Future discoveries will undoubtedly show why smiles and other signals of emotion are not faked often enough to be ignored.

This is a marvellous book, written in a lucid style. A must-read for those in the field, it is also a cornucopia of new thinking for anyone interested in evolution and behaviour. ■

Manfred Milinski is a director of the Max Planck Institute for Evolutionary Biology in Plön, Germany.
e-mail: milinski@evolbio.mpg.de

J. FETTER-VORM



Trinity conveys the power and paradox of the first atomic bomb.

Q&A Jonathan Fetter-Vorm

The graphic historian

New York-based author and illustrator Jonathan Fetter-Vorm makes hand-printed books on Darwin and other historical scientific subjects — including a hand-stitched treatise on surgical suturing. As he releases *Trinity*, a graphic history of the atomic bomb, Fetter-Vorm talks about crafting science chronicles.

NICK HIGGINS



Trinity: A Graphic History of the First Atomic Bomb

JONATHAN FETTER-VORM
Hall and Wang: 2012.
160 pp. \$22

How did you become interested in the history of science?

Growing up I thought that I was going to be a biologist, but my interest in the history of ideas took over. I did research into the invention of movable type, and how the printing press brought together

scientists, poets and illustrators during the Renaissance.

How did that lead you to print your own illustrated editions of science books from that time?

I like that period because self-taught people such as Leonardo da Vinci and Jesuit scholar Athanasius Kircher were defining what science meant. There wasn't a strong distinction between the wondrous and the demonstrably true. That informs how I approach illustration and bookmaking. By reminding readers that these theories that we hold to be objective laws of nature were

once just someone's hunch, I want to make science uncanny again.

How did you go about that with Darwin?

Rereading his books, I was struck by the imagery that Darwin had to invent to explain his ideas. My 2010 book *Incidental Sources with Subsequent Modifications* is an illustrated adaptation that lets fragments from *On the Origin of Species* (John Murray, 1859) mutate page by page in increasingly imaginative alterations and revisions.

And your book on sutures?

Early suture technique borrowed heavily from the craft of tailors. Joseph Lister, pioneer of antiseptic surgery, devised innovative and complex sutures using felt, wire and lead buttons. With *The Technic of Suture* (2009), I traced the echoes of this invention through various fields, from medicine to geology. I created an edition of five books, screen-printed onto felt that was treated with rabbit-skin glue and bound with suture stitches.

How did you approach writing *Trinity*?

My knowledge of the Manhattan Project was pretty sketchy and I knew nothing of the science of the atomic bomb, but I have always been intrigued by J. Robert Oppenheimer, one of the leaders of the project. He was such an enigmatic and ambivalent character: his opinions about the development and use of the atomic bomb were paradoxical and tortured. I wanted that sense of ambivalence to run through the story.

What difficulties did you encounter?

Showing what the inside of an atomic bomb looked like required almost as much imagination as drawing the inside of an atom. The government kept the structure of these devices secret. Even now, there is speculation about their inner workings. But the detonations were the hardest to draw; there are three in the book. For the Trinity test, the challenge was to compete with the beautiful, bizarre high-speed photos that the government took. I decided to use the Hiroshima detonation to depict damage to structures and not people. Nagasaki was the most difficult: days before the deadline, I was still struggling to draw something that one shudders even to imagine.

What should a good science illustrator do?

I want my illustrations to clarify what the readers may already know, and give them a map for what they might not yet know. Comics can serve as a visual shorthand for anything that is — like a fission reaction — invisible or incoherent. Something happens when the words and drawings come together: they can inspire wonder and, hopefully, a desire to learn more. ■

INTERVIEW BY JASCHA HOFFMAN

Correspondence

Categorize probiotics to speed research

As chairman of a panel convened in 2001–02 by the United Nations and the World Health Organization to establish guidelines for the use of probiotics, I am dismayed by recent regulatory changes in Europe and the United States.

These have led the European Food Safety Authority (EFSA) to override peer-reviewed studies in prestigious journals, and the US Food and Drug Administration (FDA) now requires an Investigational New Drug Application to be filed for probiotic foods or supplements that are intended to prevent or treat disease. The former has led to many negative ramifications for probiotic research and development in Europe, and the latter has halted research on probiotics at the US National Institutes of Health.

With the market for probiotics already exceeding US\$30 billion, and the Human Microbiome Project uncovering new candidate probiotic strains, I propose that the major agencies (at least EFSA and the FDA), together with international probiotic experts, create a simple categorization ‘tree’ system that would improve the adjudication process and be more informative for consumers. Every probiotic would need a set of minimal requirements, including strain designation and shelf life.

The lowest category would include yogurts that reduce adverse effects in lactose-intolerant individuals; other ‘category 1’ products would require only minimally documented studies in humans.

For the middle category, at least two randomized controlled studies would be needed to show how the probiotic works, with the results published in peer-reviewed journals. This could include probiotic yogurts that, in addition to the lactose-digestion

effect, reduce gut discomfort or increase intestinal transit time, as well as other probiotic products that counter adverse effects of antibiotics or restore microbial homeostasis to the vagina or mouth.

The third category would be reserved for products targeting vulnerable people such as infants and the elderly. It would include recombinant strains and species not previously used in foods and supplements — for example, bacteria producing neurochemicals that could improve cognitive function or memory. Strict adjudication would be required for products in category 3.

Gregor Reid *Lawson Health Research Institute and University of Western Ontario, London, Ontario, Canada.*
gregor@uwo.ca

Support home-grown plant collectors

Plant collection continues today, but not always for conservation or taxonomy purposes (*Nature* **484**, 436–438; 2012). The importation of ornamental plants has become a multibillion-dollar industry.

Unfortunately, some of the introduced plants destined for horticultural development can be a source of invasive alien weeds that affect human health, the economy and biodiversity. This risk can be offset by greater representation of native species in living collections and in horticulture.

We should be building national botanical knowledge and institutions to support a new generation of home-grown plant collectors in floristically rich regions. This will help to secure threatened flora and encourage botanic gardens to concentrate on the value and beauty of native species, rather than on showy introductions.

Little of the profit from the ornamental plant trade returns to the regions where the species

were collected. Many countries now restrict the unregulated movement beyond their borders of native genetic resources such as seeds.

The institutions and industries that have benefited from the great foreign plant hunters of the past should not view these restrictions as an impediment, but as a reminder that they have an obligation to invest in the resident plant collectors of the future.

Philip E. Hulme *The Bio-Protection Research Centre, Lincoln University, Canterbury, New Zealand.*
philip.hulme@lincoln.ac.nz

Careless linking of Wallace and Darwin

In his review of *Darwin's Ghosts* (*Nature* **485**, 171–172; 2012), Andrew Berry misleadingly writes, “Even Alfred Russel Wallace, co-author of the paper that first unveiled evolution by natural selection, has mostly disappeared from view”, adding, “The outcomes [of an intervention by colleagues] were a paper co-published by Darwin and Wallace in the *Journal of the Linnean Society* in July 1858, and *Origin* in November the next year.”

The terms ‘co-author’ and ‘co-publish’ wrongly imply that Darwin and Wallace were joint authors of a single paper and joint publishers of the (incorrectly titled) journal.

The facts were well stated by geologist Charles Lyell and botanist Joseph Hooker, who on 1 July 1858 presented to the Linnean Society an essay by Wallace, together with two contributions from Darwin (extracts from a manuscript on species and an abstract of a letter to US botanist Asa Gray). Lyell and Hooker wrote: “The accompanying papers [...] all relate to the same subject, viz. the Laws which affect the Production of Varieties, Races, and Species, [and] contain the

results of the investigations of two indefatigable naturalists, Mr. Charles Darwin and Mr. Alfred Wallace.” These three ‘papers’ were first printed in 1858 (*J. Proc. Linn. Soc.* **3**, 45–62; 1858), and later in volume form in 1859.

K. Razi Naqvi *Norwegian University of Science and Technology, Trondheim, Norway.*
razi.naqvi@ntnu.no

Monitoring fungal infections in fish

Freshwater fish should be added to the list of species that are threatened by emerging fungal diseases (M. C. Fisher *et al.* *Nature* **484**, 186–194; 2012). Government agencies need to adapt their fish-monitoring programmes to establish the extent of the damage these pathogens are causing.

Freshwater fish are important for millions of people in eastern Asia, for example, but fungal diseases are spreading fast there, helped by invasive species that carry fungal pathogens (R. E. Gozlan *et al.* *Fish Fish.* **11**, 315–340; 2010).

It is essential to convince cash-strapped government agencies that experimental work can give insight into real-world epidemiology, and that a major impact of disease on fish populations could go unnoticed using current monitoring systems, which don't work for fish living in muddy waters.

Rodolphe Gozlan *Bournemouth University, Poole, Dorset, UK.*
rgozlan@bournemouth.ac.uk

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the guidelines at go.nature.com/cmchno. Alternatively, readers may comment online on anything published in *Nature* at: www.nature.com/nature.

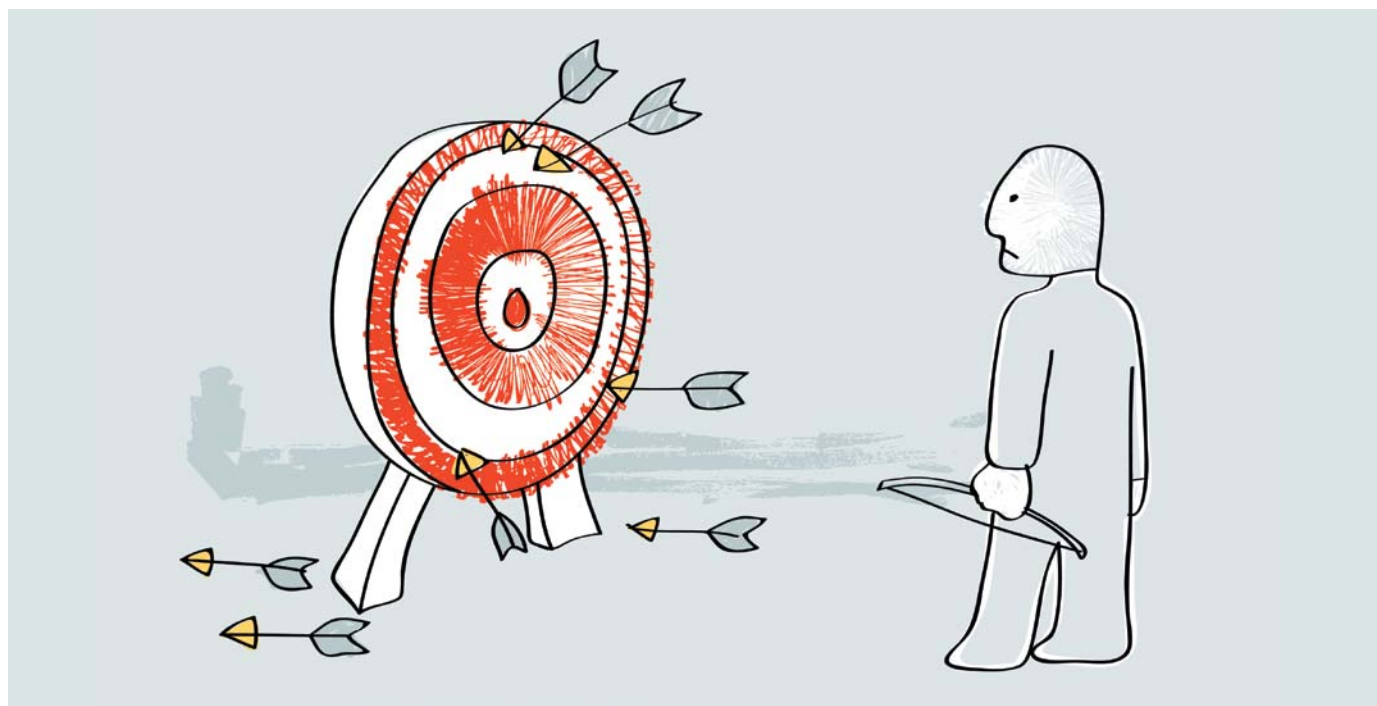
CAREERS

TURNING POINT Researcher recalls how a shift in focus led to a brain-research post **p.537**

PAY GAP Salary of Canadian female faculty falls short of their male colleagues **p.537**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

L. CHAN



POSTGRADUATE OPTIONS

Academia misses the mark

Careers advice offered at US institutes is lacking for doctoral students disillusioned with the prospect of an academic career.

BY KAREN KAPLAN

A US study confirms what many observers of and participants in graduate-level science education have long suspected: that physical and natural-sciences graduate students become less interested in academic careers as they progress through their degrees. This is the latest study to suggest that there is an increase in disillusionment as students get closer to graduation.

According to the authors of the most recent report¹, the findings underscore the importance of broad-based career guidance and mentoring at US universities, as well as the growing need for university administrators, faculty members and advisers to provide students with information about the options in sectors such as industry, government and

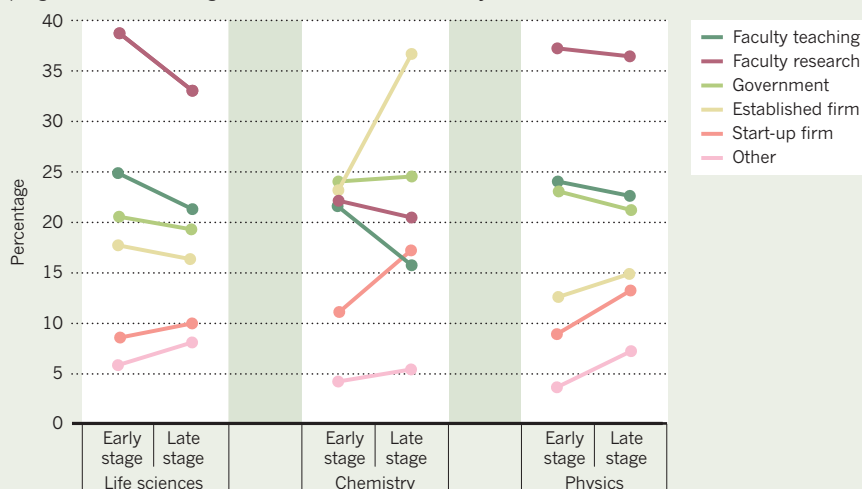
not-for-profit. It is also imperative, the authors add, for doctoral students to arm themselves with details about non-academic careers rather than relying on career-guidance services at their institutions.

The study, published on 2 May, surveyed life sciences, physics and chemistry PhD students at various stages of their programmes at top-level US research universities. Respondents in graduate programmes were asked to rate six career options and to recall how they had felt about them at the start of their PhD programme. Although a faculty post was an attractive career path for many students at the start of their programme, this preference slipped as students in all three disciplines advanced in their studies, with chemistry students showing the biggest drop (see 'Losing appeal').

A small group of respondents in all three disciplines said that faculty members, supervisors, advisers and mentors encouraged them to pursue academic research as a career option, and discouraged other career choices, including industrial research. For example, 11% of life-sciences students reported receiving a negative reaction from their laboratory supervisor or department head about pursuing a position at either an established business or a start-up company. Policy-makers, professional associations and university administrators may therefore need to supplement the career information that graduate students receive from their supervisors, and the graduate curricula should have less of a focus on academia, say the authors. Study co-author Henry Sauermann, a researcher at the College of Management at Georgia Institute of Technology in Atlanta, ►

LOSING APPEAL

By the time they got to the end of their degrees, fewer students in PhD programmes were rating academic careers as 'extremely attractive'.



SOURCE: REF. 1

► adds that their findings illustrate the disparity between the career options that advisers suggest and what positions are actually available. Sauermann encourages doctoral students to bridge this gap by researching non-academic options themselves. “It would be nice for other people to provide more information for students, but faculty and advisers don’t have that experience,” he says. “You can’t expect the chair of the chemistry department to tell you what it’s like being a researcher in industry.”

The study’s results echo those of a *Nature* survey² in June last year. A report³ co-authored by the US Council of Graduate Schools — which was compiled by university administrators and deans, industry executives and higher-education consultants, and surveyed graduate students, deans of graduate schools and employers — concluded that US universities, federal policy-makers and employers must coordinate their efforts to improve the career paths of postgraduates. “Successful career advice depends on information,” says Debra Stewart, president of the Council of Graduate Schools in Washington DC. “And one of the most important things we can do for students is to provide them with information about the career trajectory.”

Efforts to diffuse the long-term focus on academic careers at UK institutions have been under way for about a decade, says Janet Metcalfe, head of Vitae, a research-career advisory organization in Cambridge, UK. Metcalfe notes that the 2002 ‘Roberts report’ — named after its primary author, the UK scientist and education policy leader Gareth Roberts — recommended that early-career researchers receive broad-spectrum training in professional-development. As a result, she says, funds now exist to support such training at UK institutions. Vitae itself offers career guidance to junior scientists, and most universities have specialist careers advisers, Metcalfe adds.

Some top-tier US research universities have started to broaden the advice they give to science doctoral students, although there is a lingering pro-academic bias on the part of some faculty members. “Some professors probably think, ‘You don’t need to go down that route,’” says Patricia Simpson, director of the Career Services Network in the School of Chemical Sciences at the University of Illinois at Urbana-Champaign. “But most are at least open to having their students talk with me.” She says that she counsels up to 175 graduate students a year on non-academic career options. Yuree Soh, an assistant director of the Careers Development Center at Stanford University, California, emphasizes that institutional career services are crucial. She has launched several programmes, including career fairs and alumni panels, at which students can network with alumni who are not in academic posts.

Many faculty members still don’t realize that only a tiny fraction of US postgraduates land a tenure-track academic research position, says Patrick Brandt, director of science, training and diversity at the University of North Carolina in Chapel Hill. Brandt, a former biochemist, has also initiated several career-advice programmes, including one in which former researchers talk about their non-academic careers. “We are ethically bound to provide broad-spectrum career guidance to rising biomedical scientists,” he says.

Soh says that those who are in a position to influence students are showing subtle, but encouraging, signs of a shift away from the pro-academia bias. “It’s not a tidal wave,” she says. “But more questions are being asked and there is more momentum building.” ■

1. Sauermann, H. & Roach, M. *PLoS ONE* **7**, e36307 (2012).
2. Russo, G. *Nature* **475**, 533–535 (2011).
3. Wendler, C. et al. *Pathways Through Graduate School and Into Careers*. (Educational Testing Service, 2012).

TURNING POINT

Ricardo Dolmetsch

Ricardo Dolmetsch is a neurobiologist at Stanford University, California. In March he announced his affiliation with the Allen Institute for Brain Science in Seattle, Washington, where he will help to develop tools for identifying the genetic mutations that alter brain development and behaviour.

Did you always know you would be a scientist?

No. I grew up in Colombia in a family of artists. I became enamoured with science in a bizarre way: I was suspended from school for writing erotic poems in typing class and, made to stay at home for a week, I learned how to cook. I realized then that I was good at following instructions and doing manual tasks, which is a big part of being a scientist.

What piqued your interest in neurobiology?

When I decided I wanted to go to university in the United States, I knew very little about the universities there. I got accepted to Brown University in Providence, Rhode Island. Alongside my studies, I worked as a laboratory assistant for a clinician who was the head of child neurology at Rhode Island Hospital in Providence. She had lots of equipment, a technician and me. It was great. I got to work with microscopes, oscilloscopes and amplifiers, and this made me realize that I wanted to go to graduate school.

You had a difficult start to your PhD programme. What happened?

I applied to Stanford University, in part because I enjoyed my interview with the neurobiologist Carla Shatz during the pre-application process. But, as soon as I got there, she left. I joined another lab, but the principal investigator there, neuroscientist Dennis Choi, left four months later. I thought I was the kiss of death for faculty at Stanford. I was not born in the United States, so, at the time, I was not eligible for US National Institutes of Health (NIH) funding. I had to either find a new lab and funding quickly, or go back to Colombia and join the army, which, in those days, was not a fabulous option. A faculty member at Stanford suggested that I find a supervisor who was just starting their lab. I found molecular biologist Richard Lewis. Although Lewis was not a neuroscientist, he was exploring how calcium channels regulate immune function. I became his first student and, after four years and two failures, I made a machine that could control intracellular calcium patterns in cells, which led to several high-profile papers.



What was your experience of being a new faculty member at Stanford?

It can be difficult for young professors to begin with. Suddenly there are all these things to do so that you can execute your good ideas — managing and motivating people, or looking after the finances. Initially, things went pretty well. But the lab started running out of money and the situation became desperate. Then my son was diagnosed with autism, which proved to be pivotal. For the first time, I made a conscious decision to focus my career on neurodevelopmental diseases — genetic diseases, including autism, that affect cognition. To support the change in my research focus, I began to apply for big, competitive grants and, luckily, I got an NIH Pioneer award. This rescued the lab and has allowed me to stay in this area of work.

Is the position at the Allen Institute motivated by aspirations to set up a lab in Colombia?

That is one part of it. I hope to learn how to secure sustainable funding, manage people and direct a big project. I feel I owe something to the community back home that invested in my success — so, yes, at some point, I would love to start a research institute in Colombia. But until more research funding becomes available there, it is not realistic. For now, I will retain a lab at Stanford as a way to use the tools developed at the Allen Institute. ■

INTERVIEW BY VIRGINIA GEWIN

CORRECTION

The News item 'Career-path support' (*Nature* **485**, 139; 2012) wrongly located the Educational Testing Service in Washington DC instead of Princeton, New Jersey.

PAY GAP

Reaching higher ranks

Full-time male Canadian faculty members earned almost 13% more than their female counterparts during the 2010–11 academic year, according to a 3 May report by Statistics Canada, a federal agency in Ontario. Women earned, on average, Can\$106,970 (US\$105,921) in 2010–11 compared with an average of \$120,378 for men, the report says. For full professors, the report notes that females earned 5% less than males that year. Women have been slower to reach the highest academic ranks, and have not yet attained the salary levels of their longer-tenured male faculty members, says sociologist Michael Smith at McGill University in Montreal, Canada, who studies labour market and economic issues. Women account for just 23.4% of full professors, but 46.4% of assistant professors.

EDUCATION

Best global provider

The US higher-education system tops a list of 48 countries in a ranking released on 10 May. Authors compiled the list from data on resources, output, connectivity and environment for Universitas 21, an international network of research universities based in Birmingham, UK. Co-author Ross Williams, professorial fellow at the Melbourne Institute of Applied Economic and Social Research in Australia, says that the list can guide early-career researchers to the higher-ranking nations that are more likely to provide the funding, positive environment and peer interaction needed for career advancement.

NEUROSCIENCE

Brain research boost

A six-year, Can\$100-million (US\$99-million) commitment from the Canadian federal government to fund neuroscience research will include training awards for early-career neuroscientists. The Canada Brain Research Fund will award a number of postdoctoral fellowships at \$50,000 per year and PhD studentships at \$25,000 per year, which will also be open to international students who are accepted at a Canadian university. The Brain Canada Foundation in Montreal, a national charity, will administer the awards. The foundation will aim to match the government funding through private donors and organizations. The number of postdoc fellowships and studentships will be determined once matching funds are tallied.

TURNING POINT

Ricardo Dolmetsch

Ricardo Dolmetsch is a neurobiologist at Stanford University, California. In March he announced his affiliation with the Allen Institute for Brain Science in Seattle, Washington, where he will help to develop tools for identifying the genetic mutations that alter brain development and behaviour.

Did you always know you would be a scientist?

No. I grew up in Colombia in a family of artists. I became enamoured with science in a bizarre way: I was suspended from school for writing erotic poems in typing class and, made to stay at home for a week, I learned how to cook. I realized then that I was good at following instructions and doing manual tasks, which is a big part of being a scientist.

What piqued your interest in neurobiology?

When I decided I wanted to go to university in the United States, I knew very little about the universities there. I got accepted to Brown University in Providence, Rhode Island. Alongside my studies, I worked as a laboratory assistant for a clinician who was the head of child neurology at Rhode Island Hospital in Providence. She had lots of equipment, a technician and me. It was great. I got to work with microscopes, oscilloscopes and amplifiers, and this made me realize that I wanted to go to graduate school.

You had a difficult start to your PhD programme. What happened?

I applied to Stanford University, in part because I enjoyed my interview with the neurobiologist Carla Shatz during the pre-application process. But, as soon as I got there, she left. I joined another lab, but the principal investigator there, neuroscientist Dennis Choi, left four months later. I thought I was the kiss of death for faculty at Stanford. I was not born in the United States, so, at the time, I was not eligible for US National Institutes of Health (NIH) funding. I had to either find a new lab and funding quickly, or go back to Colombia and join the army, which, in those days, was not a fabulous option. A faculty member at Stanford suggested that I find a supervisor who was just starting their lab. I found molecular biologist Richard Lewis. Although Lewis was not a neuroscientist, he was exploring how calcium channels regulate immune function. I became his first student and, after four years and two failures, I made a machine that could control intracellular calcium patterns in cells, which led to several high-profile papers.



What was your experience of being a new faculty member at Stanford?

It can be difficult for young professors to begin with. Suddenly there are all these things to do so that you can execute your good ideas — managing and motivating people, or looking after the finances. Initially, things went pretty well. But the lab started running out of money and the situation became desperate. Then my son was diagnosed with autism, which proved to be pivotal. For the first time, I made a conscious decision to focus my career on neurodevelopmental diseases — genetic diseases, including autism, that affect cognition. To support the change in my research focus, I began to apply for big, competitive grants and, luckily, I got an NIH Pioneer award. This rescued the lab and has allowed me to stay in this area of work.

Is the position at the Allen Institute motivated by aspirations to set up a lab in Colombia?

That is one part of it. I hope to learn how to secure sustainable funding, manage people and direct a big project. I feel I owe something to the community back home that invested in my success — so, yes, at some point, I would love to start a research institute in Colombia. But until more research funding becomes available there, it is not realistic. For now, I will retain a lab at Stanford as a way to use the tools developed at the Allen Institute. ■

INTERVIEW BY VIRGINIA GEWIN

CORRECTION

The News item 'Career-path support' (*Nature* **485**, 139; 2012) wrongly located the Educational Testing Service in Washington DC instead of Princeton, New Jersey.

PAY GAP

Reaching higher ranks

Full-time male Canadian faculty members earned almost 13% more than their female counterparts during the 2010–11 academic year, according to a 3 May report by Statistics Canada, a federal agency in Ontario. Women earned, on average, Can\$106,970 (US\$105,921) in 2010–11 compared with an average of \$120,378 for men, the report says. For full professors, the report notes that females earned 5% less than males that year. Women have been slower to reach the highest academic ranks, and have not yet attained the salary levels of their longer-tenured male faculty members, says sociologist Michael Smith at McGill University in Montreal, Canada, who studies labour market and economic issues. Women account for just 23.4% of full professors, but 46.4% of assistant professors.

EDUCATION

Best global provider

The US higher-education system tops a list of 48 countries in a ranking released on 10 May. Authors compiled the list from data on resources, output, connectivity and environment for Universitas 21, an international network of research universities based in Birmingham, UK. Co-author Ross Williams, professorial fellow at the Melbourne Institute of Applied Economic and Social Research in Australia, says that the list can guide early-career researchers to the higher-ranking nations that are more likely to provide the funding, positive environment and peer interaction needed for career advancement.

NEUROSCIENCE

Brain research boost

A six-year, Can\$100-million (US\$99-million) commitment from the Canadian federal government to fund neuroscience research will include training awards for early-career neuroscientists. The Canada Brain Research Fund will award a number of postdoctoral fellowships at \$50,000 per year and PhD studentships at \$25,000 per year, which will also be open to international students who are accepted at a Canadian university. The Brain Canada Foundation in Montreal, a national charity, will administer the awards. The foundation will aim to match the government funding through private donors and organizations. The number of postdoc fellowships and studentships will be determined once matching funds are tallied.

TURNING POINT

Ricardo Dolmetsch

Ricardo Dolmetsch is a neurobiologist at Stanford University, California. In March he announced his affiliation with the Allen Institute for Brain Science in Seattle, Washington, where he will help to develop tools for identifying the genetic mutations that alter brain development and behaviour.

Did you always know you would be a scientist?

No. I grew up in Colombia in a family of artists. I became enamoured with science in a bizarre way: I was suspended from school for writing erotic poems in typing class and, made to stay at home for a week, I learned how to cook. I realized then that I was good at following instructions and doing manual tasks, which is a big part of being a scientist.

What piqued your interest in neurobiology?

When I decided I wanted to go to university in the United States, I knew very little about the universities there. I got accepted to Brown University in Providence, Rhode Island. Alongside my studies, I worked as a laboratory assistant for a clinician who was the head of child neurology at Rhode Island Hospital in Providence. She had lots of equipment, a technician and me. It was great. I got to work with microscopes, oscilloscopes and amplifiers, and this made me realize that I wanted to go to graduate school.

You had a difficult start to your PhD programme. What happened?

I applied to Stanford University, in part because I enjoyed my interview with the neurobiologist Carla Shatz during the pre-application process. But, as soon as I got there, she left. I joined another lab, but the principal investigator there, neuroscientist Dennis Choi, left four months later. I thought I was the kiss of death for faculty at Stanford. I was not born in the United States, so, at the time, I was not eligible for US National Institutes of Health (NIH) funding. I had to either find a new lab and funding quickly, or go back to Colombia and join the army, which, in those days, was not a fabulous option. A faculty member at Stanford suggested that I find a supervisor who was just starting their lab. I found molecular biologist Richard Lewis. Although Lewis was not a neuroscientist, he was exploring how calcium channels regulate immune function. I became his first student and, after four years and two failures, I made a machine that could control intracellular calcium patterns in cells, which led to several high-profile papers.



What was your experience of being a new faculty member at Stanford?

It can be difficult for young professors to begin with. Suddenly there are all these things to do so that you can execute your good ideas — managing and motivating people, or looking after the finances. Initially, things went pretty well. But the lab started running out of money and the situation became desperate. Then my son was diagnosed with autism, which proved to be pivotal. For the first time, I made a conscious decision to focus my career on neurodevelopmental diseases — genetic diseases, including autism, that affect cognition. To support the change in my research focus, I began to apply for big, competitive grants and, luckily, I got an NIH Pioneer award. This rescued the lab and has allowed me to stay in this area of work.

Is the position at the Allen Institute motivated by aspirations to set up a lab in Colombia?

That is one part of it. I hope to learn how to secure sustainable funding, manage people and direct a big project. I feel I owe something to the community back home that invested in my success — so, yes, at some point, I would love to start a research institute in Colombia. But until more research funding becomes available there, it is not realistic. For now, I will retain a lab at Stanford as a way to use the tools developed at the Allen Institute. ■

INTERVIEW BY VIRGINIA GEWIN

CORRECTION

The News item 'Career-path support' (*Nature* **485**, 139; 2012) wrongly located the Educational Testing Service in Washington DC instead of Princeton, New Jersey.

PAY GAP

Reaching higher ranks

Full-time male Canadian faculty members earned almost 13% more than their female counterparts during the 2010–11 academic year, according to a 3 May report by Statistics Canada, a federal agency in Ontario. Women earned, on average, Can\$106,970 (US\$105,921) in 2010–11 compared with an average of \$120,378 for men, the report says. For full professors, the report notes that females earned 5% less than males that year. Women have been slower to reach the highest academic ranks, and have not yet attained the salary levels of their longer-tenured male faculty members, says sociologist Michael Smith at McGill University in Montreal, Canada, who studies labour market and economic issues. Women account for just 23.4% of full professors, but 46.4% of assistant professors.

EDUCATION

Best global provider

The US higher-education system tops a list of 48 countries in a ranking released on 10 May. Authors compiled the list from data on resources, output, connectivity and environment for Universitas 21, an international network of research universities based in Birmingham, UK. Co-author Ross Williams, professorial fellow at the Melbourne Institute of Applied Economic and Social Research in Australia, says that the list can guide early-career researchers to the higher-ranking nations that are more likely to provide the funding, positive environment and peer interaction needed for career advancement.

NEUROSCIENCE

Brain research boost

A six-year, Can\$100-million (US\$99-million) commitment from the Canadian federal government to fund neuroscience research will include training awards for early-career neuroscientists. The Canada Brain Research Fund will award a number of postdoctoral fellowships at \$50,000 per year and PhD studentships at \$25,000 per year, which will also be open to international students who are accepted at a Canadian university. The Brain Canada Foundation in Montreal, a national charity, will administer the awards. The foundation will aim to match the government funding through private donors and organizations. The number of postdoc fellowships and studentships will be determined once matching funds are tallied.

TURNING POINT

Ricardo Dolmetsch

Ricardo Dolmetsch is a neurobiologist at Stanford University, California. In March he announced his affiliation with the Allen Institute for Brain Science in Seattle, Washington, where he will help to develop tools for identifying the genetic mutations that alter brain development and behaviour.

Did you always know you would be a scientist?

No. I grew up in Colombia in a family of artists. I became enamoured with science in a bizarre way: I was suspended from school for writing erotic poems in typing class and, made to stay at home for a week, I learned how to cook. I realized then that I was good at following instructions and doing manual tasks, which is a big part of being a scientist.

What piqued your interest in neurobiology?

When I decided I wanted to go to university in the United States, I knew very little about the universities there. I got accepted to Brown University in Providence, Rhode Island. Alongside my studies, I worked as a laboratory assistant for a clinician who was the head of child neurology at Rhode Island Hospital in Providence. She had lots of equipment, a technician and me. It was great. I got to work with microscopes, oscilloscopes and amplifiers, and this made me realize that I wanted to go to graduate school.

You had a difficult start to your PhD programme. What happened?

I applied to Stanford University, in part because I enjoyed my interview with the neurobiologist Carla Shatz during the pre-application process. But, as soon as I got there, she left. I joined another lab, but the principal investigator there, neuroscientist Dennis Choi, left four months later. I thought I was the kiss of death for faculty at Stanford. I was not born in the United States, so, at the time, I was not eligible for US National Institutes of Health (NIH) funding. I had to either find a new lab and funding quickly, or go back to Colombia and join the army, which, in those days, was not a fabulous option. A faculty member at Stanford suggested that I find a supervisor who was just starting their lab. I found molecular biologist Richard Lewis. Although Lewis was not a neuroscientist, he was exploring how calcium channels regulate immune function. I became his first student and, after four years and two failures, I made a machine that could control intracellular calcium patterns in cells, which led to several high-profile papers.



What was your experience of being a new faculty member at Stanford?

It can be difficult for young professors to begin with. Suddenly there are all these things to do so that you can execute your good ideas — managing and motivating people, or looking after the finances. Initially, things went pretty well. But the lab started running out of money and the situation became desperate. Then my son was diagnosed with autism, which proved to be pivotal. For the first time, I made a conscious decision to focus my career on neurodevelopmental diseases — genetic diseases, including autism, that affect cognition. To support the change in my research focus, I began to apply for big, competitive grants and, luckily, I got an NIH Pioneer award. This rescued the lab and has allowed me to stay in this area of work.

Is the position at the Allen Institute motivated by aspirations to set up a lab in Colombia?

That is one part of it. I hope to learn how to secure sustainable funding, manage people and direct a big project. I feel I owe something to the community back home that invested in my success — so, yes, at some point, I would love to start a research institute in Colombia. But until more research funding becomes available there, it is not realistic. For now, I will retain a lab at Stanford as a way to use the tools developed at the Allen Institute. ■

INTERVIEW BY VIRGINIA GEWIN

CORRECTION

The News item 'Career-path support' (*Nature* **485**, 139; 2012) wrongly located the Educational Testing Service in Washington DC instead of Princeton, New Jersey.

PAY GAP

Reaching higher ranks

Full-time male Canadian faculty members earned almost 13% more than their female counterparts during the 2010–11 academic year, according to a 3 May report by Statistics Canada, a federal agency in Ontario. Women earned, on average, Can\$106,970 (US\$105,921) in 2010–11 compared with an average of \$120,378 for men, the report says. For full professors, the report notes that females earned 5% less than males that year. Women have been slower to reach the highest academic ranks, and have not yet attained the salary levels of their longer-tenured male faculty members, says sociologist Michael Smith at McGill University in Montreal, Canada, who studies labour market and economic issues. Women account for just 23.4% of full professors, but 46.4% of assistant professors.

EDUCATION

Best global provider

The US higher-education system tops a list of 48 countries in a ranking released on 10 May. Authors compiled the list from data on resources, output, connectivity and environment for Universitas 21, an international network of research universities based in Birmingham, UK. Co-author Ross Williams, professorial fellow at the Melbourne Institute of Applied Economic and Social Research in Australia, says that the list can guide early-career researchers to the higher-ranking nations that are more likely to provide the funding, positive environment and peer interaction needed for career advancement.

NEUROSCIENCE

Brain research boost

A six-year, Can\$100-million (US\$99-million) commitment from the Canadian federal government to fund neuroscience research will include training awards for early-career neuroscientists. The Canada Brain Research Fund will award a number of postdoctoral fellowships at \$50,000 per year and PhD studentships at \$25,000 per year, which will also be open to international students who are accepted at a Canadian university. The Brain Canada Foundation in Montreal, a national charity, will administer the awards. The foundation will aim to match the government funding through private donors and organizations. The number of postdoc fellowships and studentships will be determined once matching funds are tallied.

TURNING POINT

Ricardo Dolmetsch

Ricardo Dolmetsch is a neurobiologist at Stanford University, California. In March he announced his affiliation with the Allen Institute for Brain Science in Seattle, Washington, where he will help to develop tools for identifying the genetic mutations that alter brain development and behaviour.

Did you always know you would be a scientist?

No. I grew up in Colombia in a family of artists. I became enamoured with science in a bizarre way: I was suspended from school for writing erotic poems in typing class and, made to stay at home for a week, I learned how to cook. I realized then that I was good at following instructions and doing manual tasks, which is a big part of being a scientist.

What piqued your interest in neurobiology?

When I decided I wanted to go to university in the United States, I knew very little about the universities there. I got accepted to Brown University in Providence, Rhode Island. Alongside my studies, I worked as a laboratory assistant for a clinician who was the head of child neurology at Rhode Island Hospital in Providence. She had lots of equipment, a technician and me. It was great. I got to work with microscopes, oscilloscopes and amplifiers, and this made me realize that I wanted to go to graduate school.

You had a difficult start to your PhD programme. What happened?

I applied to Stanford University, in part because I enjoyed my interview with the neurobiologist Carla Shatz during the pre-application process. But, as soon as I got there, she left. I joined another lab, but the principal investigator there, neuroscientist Dennis Choi, left four months later. I thought I was the kiss of death for faculty at Stanford. I was not born in the United States, so, at the time, I was not eligible for US National Institutes of Health (NIH) funding. I had to either find a new lab and funding quickly, or go back to Colombia and join the army, which, in those days, was not a fabulous option. A faculty member at Stanford suggested that I find a supervisor who was just starting their lab. I found molecular biologist Richard Lewis. Although Lewis was not a neuroscientist, he was exploring how calcium channels regulate immune function. I became his first student and, after four years and two failures, I made a machine that could control intracellular calcium patterns in cells, which led to several high-profile papers.



What was your experience of being a new faculty member at Stanford?

It can be difficult for young professors to begin with. Suddenly there are all these things to do so that you can execute your good ideas — managing and motivating people, or looking after the finances. Initially, things went pretty well. But the lab started running out of money and the situation became desperate. Then my son was diagnosed with autism, which proved to be pivotal. For the first time, I made a conscious decision to focus my career on neurodevelopmental diseases — genetic diseases, including autism, that affect cognition. To support the change in my research focus, I began to apply for big, competitive grants and, luckily, I got an NIH Pioneer award. This rescued the lab and has allowed me to stay in this area of work.

Is the position at the Allen Institute motivated by aspirations to set up a lab in Colombia?

That is one part of it. I hope to learn how to secure sustainable funding, manage people and direct a big project. I feel I owe something to the community back home that invested in my success — so, yes, at some point, I would love to start a research institute in Colombia. But until more research funding becomes available there, it is not realistic. For now, I will retain a lab at Stanford as a way to use the tools developed at the Allen Institute. ■

INTERVIEW BY VIRGINIA GEWIN

CORRECTION

The News item 'Career-path support' (*Nature* **485**, 139; 2012) wrongly located the Educational Testing Service in Washington DC instead of Princeton, New Jersey.

PAY GAP

Reaching higher ranks

Full-time male Canadian faculty members earned almost 13% more than their female counterparts during the 2010–11 academic year, according to a 3 May report by Statistics Canada, a federal agency in Ontario. Women earned, on average, Can\$106,970 (US\$105,921) in 2010–11 compared with an average of \$120,378 for men, the report says. For full professors, the report notes that females earned 5% less than males that year. Women have been slower to reach the highest academic ranks, and have not yet attained the salary levels of their longer-tenured male faculty members, says sociologist Michael Smith at McGill University in Montreal, Canada, who studies labour market and economic issues. Women account for just 23.4% of full professors, but 46.4% of assistant professors.

EDUCATION

Best global provider

The US higher-education system tops a list of 48 countries in a ranking released on 10 May. Authors compiled the list from data on resources, output, connectivity and environment for Universitas 21, an international network of research universities based in Birmingham, UK. Co-author Ross Williams, professorial fellow at the Melbourne Institute of Applied Economic and Social Research in Australia, says that the list can guide early-career researchers to the higher-ranking nations that are more likely to provide the funding, positive environment and peer interaction needed for career advancement.

NEUROSCIENCE

Brain research boost

A six-year, Can\$100-million (US\$99-million) commitment from the Canadian federal government to fund neuroscience research will include training awards for early-career neuroscientists. The Canada Brain Research Fund will award a number of postdoctoral fellowships at \$50,000 per year and PhD studentships at \$25,000 per year, which will also be open to international students who are accepted at a Canadian university. The Brain Canada Foundation in Montreal, a national charity, will administer the awards. The foundation will aim to match the government funding through private donors and organizations. The number of postdoc fellowships and studentships will be determined once matching funds are tallied.

New evidence on testosterone and cooperation

ARISING FROM C. Eisenegger, M. Naef, R. Snozzi, M. Heinrichs & E. Fehr *Nature* **463**, 356–359 (2010)

In February 2010, Eisenegger *et al.* reported increased fair bargaining behaviour after administration of testosterone in an ultimatum game¹. However, unfair offers in the ultimatum game typically are rejected; thus, not only the motives for social cooperation but also the threat of financial punishment may have accounted for these effects. Here, using the public goods game (PGG), we unambiguously show increased social cooperation after testosterone administration, but only among subjects with low levels of prenatal testosterone (measured by the right

hand's second-to-fourth-digit ratio (2D:4D)). This finding establishes positive effects of testosterone on social cooperation, with prenatal hormonal priming providing for important individual variability.

Eisenegger *et al.* show increased fairness in bargaining behaviour after testosterone administration in young females, and the authors suggest that this prosocial behaviour is strategically driven by concerns for social status¹. Indeed, in the ultimatum game such strategic concerns have a role, and the hormone testosterone repeatedly has been associated with status concerns in humans and other animals^{2,3}. However, unfair ultimatum game offers are typically rejected with all money being lost. Hence, the threat of financial punishment may have played a part in fair bargaining behaviour after testosterone administration¹. We therefore tested the effects of testosterone on social-cooperative behaviours with the PGG, a game without such threat of financial punishment, wherein non-cooperation can actually lead to greater profits⁴. In an experiment (approved by our ethics committee) we administered testosterone and placebo on separate days to 24 female students in a double-blind within-subject design⁵, and tested them in a three-player PGG lasting eight rounds. Each round the players received an endowment of 3 monetary units (MU), which they could either keep for themselves or contribute to the public good⁴ (see Methods).

Using a repeated-measures generalized estimating equations (GEE) analysis over all eight trials (PGG placebo versus testosterone) we found no main effect of testosterone (Wald $\chi^2 = 0.048$, $P = 0.826$). However, we also measured a proxy of prenatal testosterone, the 2D:4D ratio^{6,7}, which has recently been shown to be a powerful predictor for effects of testosterone administration on social function⁸ (Methods). With 2D:4D ratio as covariate in the analyses the effect of testosterone on social cooperation was significant (Wald $\chi^2 = 9.630$, $P = 0.002$) as, importantly, was the 2D:4D \times testosterone interaction (Wald $\chi^2 = 10.140$, $P = 0.001$) (Fig. 1a). Next, we applied a median split on the 2D:4D measurements to compare individuals with relatively low versus high prenatal testosterone exposure. GEE analyses computed in both groups separately showed that subjects with low prenatal testosterone exposure (high 2D:4D) contributed more to the group after testosterone administration (Wald $\chi^2 = 7.894$, $P = 0.005$), whereas subjects with high prenatal testosterone exposure (low 2D:4D) showed no change (Wald $\chi^2 = 1.791$, $P = 0.181$) (Fig. 1b). A forced-choice test establishing that subjects were unaware of treatment condition also revealed no belief effects on public good contributions (all P values > 0.10). Thus, unlike in the Eisenegger *et al.* study, folk beliefs about testosterone did not mediate behaviour in the PGG¹, which is not unexpected because unfair ultimatum game offers (and not PGG non-contributions) are antisocial and risky, and fit with mainstream ideas on how testosterone affects behaviour^{9,10}.

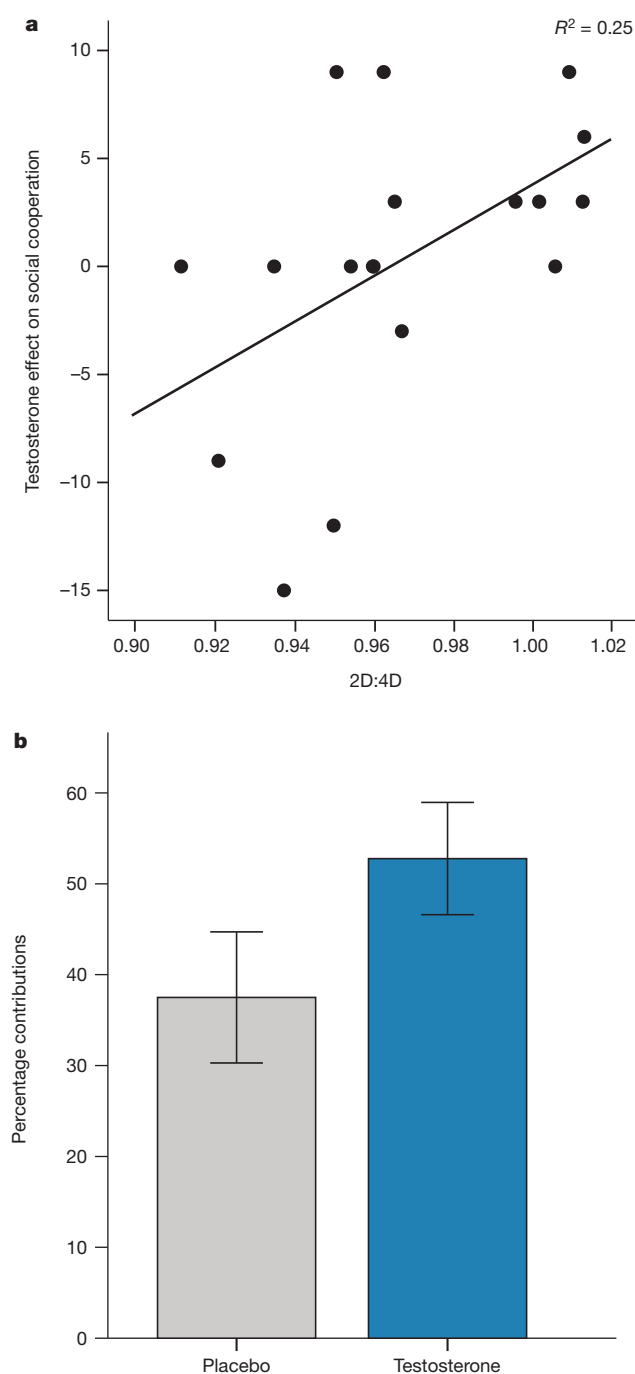


Figure 1 | Prenatal sex hormone priming (2D:4D) moderates effects of testosterone on social cooperation. **a**, Individual 2D:4D measurements plotted against testosterone effect on social cooperation (mean amount of MU contributed in the placebo condition subtracted from mean amount of MU contributed after testosterone administration). Line depicts regression wherein 2D:4D ratio explains 25% of the variance in the overall effect of testosterone on cooperation. **b**, Mean and standard error of the mean of the percentage of overall PGG contribution after testosterone and placebo administration in subjects with relatively low prenatal testosterone (high 2D:4D) on the basis of median split. The data show significantly more overall contribution to the public good after administration of testosterone compared to placebo in high 2D:4D subjects. We thank D. Tromp for her help in collecting these data.

BRIEF COMMUNICATIONS ARISING

The present result supports past research in which we also show effects of testosterone that vary strongly with prenatal testosterone exposure. In that case, high prenatal testosterone exposure (low 2D:4D) boosted the negative effect of testosterone administration on cognitive empathy⁸. Crucially, the 2D:4D ratio apparently is interactively shaped by testosterone and oestradiol *in utero*, and a high 2D:4D ratio points to relatively low prenatal testosterone versus high prenatal oestradiol^{11–13}. Furthermore, many effects of testosterone on social behaviour are thought to arise after metabolism to oestradiol^{2,3}, but this metabolism differs between individuals¹⁴. Hypothetically, the balance between the sex steroids prenatally, marked by the 2D:4D ratio^{11–13}, would be predictive for the rate of metabolism of testosterone into oestradiol. That is, subjects who are prenatally more strongly primed by oestradiol may also metabolize more testosterone into oestradiol¹³, and this could have caused the selective effect in our high 2D:4D group. Further research is necessary to test this hypothesis.

Here we challenge Eisenegger *et al.* by establishing the positive effects of testosterone on social cooperation in which prenatal sex hormone priming, approximated by determining 2D:4D ratios, conveys important individual variability. These data have strong implications for past and future hormone research.

METHODS SUMMARY

For the PGG, all three players receive 3 MU per round and can contribute all or nothing to the public good. Only when at least two players contribute does each player receive an extra 6 MU irrespective of whether they made a contribution; thus non-contributors can profit most (9 MU). To create three-person groups, confederates were used to ensure there were three players involved each time, and their decisions were randomized. After the experiment we checked that no suspicions were raised about this procedure (six non-believers were excluded from analyses). For 2D:4D measurements, subjects' right hands were scanned, and 2D:4D ratios were computed twice by an experienced rater¹⁵ (correlation between measurements, $P < 0.0001$).

Jack van Honk^{1,2}, Estrella R. Montoya¹, Peter A. Bos^{1,2}, Mark van Vugt³ & David Terburg^{1,2}

¹Department of Experimental Psychology, Utrecht University, 3584CS Utrecht, Netherlands.
email: j.vanhonk@uu.nl

²Department of Psychiatry, University of Cape Town, 7925 Cape Town, South Africa.

³Department of Social and Organizational Psychology, VU University, 1081 BT Amsterdam, Netherlands.

Received 19 October 2011; accepted 13 March 2012.

1. Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M. & Fehr, E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* **463**, 356–359 (2010).
2. Bos, P. A., Panksepp, J., Bluthé, R. M. & van Honk, J. Acute effects of steroid hormones and neuropeptides on human social-emotional behavior: a review of single administration studies. *Front. Neuroendocrinol.* **33**, 17–35 (2012).
3. Eisenegger, C., Haushofer, J. & Fehr, E. The role of testosterone in social interaction. *Trends Cogn. Sci.* **15**, 263–271 (2011).
4. Van Vugt, M., De Cremer, D. & Janssen, D. P. Gender differences in cooperation and competition: the male-warrior hypothesis. *Psychol. Sci.* **18**, 19–23 (2007).
5. Tuiten, A. *et al.* Time course of effects of testosterone administration on sexual arousal in women. *Arch. Gen. Psychiatry* **57**, 149–153 (2000).
6. Breedlove, S. M. Organizational hypothesis: instances of the fingerpost. *Endocrinology* **151**, 4116–4122 (2010).
7. Hönekopp, J., Bartholdt, L., Beier, L. & Liebert, A. Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: new data and a meta-analytic review. *Psychoneuroendocrinology* **32**, 313–321 (2007).
8. van Honk, J. *et al.* Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proc. Natl Acad. Sci. USA* **108**, 3448–3452 (2011).
9. Dabbs, J. M. & Dabbs, M. G. *Heroes, Rogues and Lovers: On Testosterone and Behavior* (McGraw-Hill, 2000).
10. van Honk, J. *et al.* Testosterone shifts the balance between sensitivity for punishment and reward in healthy young women. *Psychoneuroendocrinology* **29**, 937–943 (2004).
11. Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R. & Manning, J. T. 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Hum. Dev.* **77**, 23–28 (2004).
12. Manning, J. T., Scutt, D., Wilson, J. & Lewis-Jones, D. I. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Hum. Reprod.* **13**, 3000–3004 (1998).
13. Zheng, Z. & Cohn, M. J. Developmental basis of sexually dimorphic digit ratios. *Proc. Natl Acad. Sci. USA* **108**, 16289–16294 (2011).
14. Sarachana, T., Xu, M., Wu, R. C. & Hu, V. W. Sex hormones in autism: androgens and estrogens differentially and reciprocally regulate RORA, a novel candidate gene for autism. *PLoS ONE* **6**, e17116 (2011).
15. Millet, K. & Dewitte, S. Second to fourth digit ratio and cooperative behavior. *Biol. Psychol.* **71**, 111–115 (2006).

Author Contributions All authors were involved in the design of the research; D.T. and E.R.M. analysed data; and all authors were involved in the writing of the paper.

Competing Financial Interests Declared none.

doi:10.1038/nature11136

Eisenegger *et al.* reply

REPLYING TO J. van Honk, E. R. Montoya, P. A. Bos, M. van Vugt & D. Terburg *Nature* **485**, <http://dx.doi.org/10.1038/nature11136> (2012)

van Honk and colleagues have taken our findings on the role of testosterone in ultimatum game bargaining¹ a step forward by showing that the hormone has important prosocial effects beyond the ultimatum game by increasing cooperation in the public goods game (PGG)². In contrast to the ultimatum game, participants in the PGG decide simultaneously about their cooperation levels and are not confronted with a rejection threat from other participants, suggesting a much more universal effect of testosterone on prosociality than revealed by our study¹. As the PGG captures a large class of evolutionarily and contemporaneously important situations, their findings are of great interest, and considerably extend our knowledge about the causal effect of testosterone on social behaviour. In addition, their results raise intriguing questions regarding the motivational and biological mechanisms through which testosterone increased cooperation levels, suggesting that the study will trigger further important experiments.

One possibility is that testosterone directly affects people's social preferences, that is, it directly renders their motives more prosocial—an explanation that seems to be favoured by van Honk and colleagues². A second possibility is that the hormone influences subjects' beliefs about the other players' cooperation levels. For example, a subject who received testosterone may believe that others are more cooperative while their own motives remain unchanged. The rules of the PGG used in van Honk *et al.* are such that it is in the third player's financial interest to contribute to the public good if that subject believes that precisely one of the two other players contributes her endowment as well. Thus it is possible that the positive effect of testosterone on cooperativeness is due to the hormone's effect on beliefs about other participants' cooperation level. Of course, the two mechanisms described above are not mutually exclusive; testosterone may have causally affected subjects' motives and beliefs about others' cooperation. So far, no study has examined testosterone's causal effects

BRIEF COMMUNICATIONS ARISING

on beliefs about others' social preferences or others' prosocial behaviour—suggesting this as a topic for future studies.

Interestingly, the results of van Honk *et al.* are also compatible with the hypothesis that testosterone increases concerns for one's status^{3,4}. One study⁵ has shown that participants in a PGG confer higher status to cooperative group members than to non-cooperative ones. Another study⁶ reported that third-party observers who did not benefit in any way from high contributions also rated cooperative individuals higher on status. Thus, if subjects have internalized the link between cooperation in the PGG and their status in the group, the increased cooperation levels observed by van Honk *et al.* could have been caused by an increased concern about one's status.

The study by van Honk and colleagues also raises intriguing questions regarding the biological mechanisms underlying the hormone's effect on social behaviour. As testosterone is partially converted to oestradiol by the enzyme aromatase, the results obtained by van Honk *et al.* suggest that testosterone's effect on cooperation in the PGG in subjects with a high digit ratio might be mediated by oestradiol. This idea is based on a recent study which shows that the digit ratio is related to an individual's relative amount of androgen to oestrogen activity during early fetal development, and that oestradiol heightens the digit ratio⁷. This suggests that effects of testosterone on social behaviour may be mediated by an individual's relative androgen versus oestrogen system activity or sensitivity, for which the digit ratio might serve as an important biomarker^{7,8}. Thus, future studies should test whether a high digit ratio directly predicts social cooperation after oestradiol administration, as this might provide further insights into the biological mechanism behind the behavioural impact of sex hormones.

The study by van Honk *et al.* considerably broadens the causal role of testosterone on important prosocial behaviours while simultaneously suggesting a more specific biological underpinning based on the idea that the digit ratio is a biomarker for individuals' androgen

versus oestrogen system activity or sensitivity. In addition, the study raises important questions regarding the mechanisms underlying the relationship between testosterone administration and prosocial behaviour. In particular, studies that are able to distinguish between the three hypotheses—that testosterone affects social preferences, beliefs, or status seeking—are likely to move the field forward further.

C. Eisenegger¹, M. Naef^{1,2}, R. Snozzi¹, M. Heinrichs³ & E. Fehr¹

¹Department of Economics, Laboratory for Social and Neural Systems Research, University of Zurich, 8006 Zurich, Switzerland.

email: christoph.eisenegger@gmail.com

²Department of Economics, Royal Holloway, University of London, Egham TW20 0EX, UK.

³Department of Psychology, Laboratory for Biological and Personality Psychology, University of Freiburg, 79104 Freiburg, Germany.

1. Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M. & Fehr, E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* **463**, 356–359 (2010).
2. van Honk, J., Montoya, E. R., Bos, P. A., van Vugt, M. & Terburg, D. New evidence on testosterone and cooperation. *Nature* **485**, <http://dx.doi.org/10.1038/nature11136> (2012).
3. Mazur, A. & Booth, A. Testosterone and dominance in men. *Behav. Brain Sci.* **21**, 353–397 (1998).
4. Eisenegger, C., Haushofer, J. & Fehr, E. The role of testosterone in social interaction. *Trends Cogn. Sci.* **15**, 263–271 (2011).
5. Hardy, C. L. & Van Vugt, M. Nice guys finish first: the competitive altruism hypothesis. *Pers. Soc. Psychol. Bull.* **32**, 1402–1413 (2006).
6. Willer, R. Groups reward individual sacrifice: the status solution to the collective action problem. *Am. Sociol. Rev.* **74**, 23–43 (2009).
7. Zheng, Z. & Cohn, M. J. Developmental basis of sexually dimorphic digit ratios. *Proc. Natl Acad. Sci. USA* **108**, 16289–16294 (2011).
8. van Honk, J. *et al.* Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proc. Natl Acad. Sci. USA* **108**, 3448–3452 (2011).

doi:10.1038/nature11137

New evidence on testosterone and cooperation

ARISING FROM C. Eisenegger, M. Naef, R. Snozzi, M. Heinrichs & E. Fehr *Nature* **463**, 356–359 (2010)

In February 2010, Eisenegger *et al.* reported increased fair bargaining behaviour after administration of testosterone in an ultimatum game¹. However, unfair offers in the ultimatum game typically are rejected; thus, not only the motives for social cooperation but also the threat of financial punishment may have accounted for these effects. Here, using the public goods game (PGG), we unambiguously show increased social cooperation after testosterone administration, but only among subjects with low levels of prenatal testosterone (measured by the right

hand's second-to-fourth-digit ratio (2D:4D)). This finding establishes positive effects of testosterone on social cooperation, with prenatal hormonal priming providing for important individual variability.

Eisenegger *et al.* show increased fairness in bargaining behaviour after testosterone administration in young females, and the authors suggest that this prosocial behaviour is strategically driven by concerns for social status¹. Indeed, in the ultimatum game such strategic concerns have a role, and the hormone testosterone repeatedly has been associated with status concerns in humans and other animals^{2,3}. However, unfair ultimatum game offers are typically rejected with all money being lost. Hence, the threat of financial punishment may have played a part in fair bargaining behaviour after testosterone administration¹. We therefore tested the effects of testosterone on social-cooperative behaviours with the PGG, a game without such threat of financial punishment, wherein non-cooperation can actually lead to greater profits⁴. In an experiment (approved by our ethics committee) we administered testosterone and placebo on separate days to 24 female students in a double-blind within-subject design⁵, and tested them in a three-player PGG lasting eight rounds. Each round the players received an endowment of 3 monetary units (MU), which they could either keep for themselves or contribute to the public good⁴ (see Methods).

Using a repeated-measures generalized estimating equations (GEE) analysis over all eight trials (PGG placebo versus testosterone) we found no main effect of testosterone (Wald $\chi^2 = 0.048$, $P = 0.826$). However, we also measured a proxy of prenatal testosterone, the 2D:4D ratio^{6,7}, which has recently been shown to be a powerful predictor for effects of testosterone administration on social function⁸ (Methods). With 2D:4D ratio as covariate in the analyses the effect of testosterone on social cooperation was significant (Wald $\chi^2 = 9.630$, $P = 0.002$) as, importantly, was the 2D:4D \times testosterone interaction (Wald $\chi^2 = 10.140$, $P = 0.001$) (Fig. 1a). Next, we applied a median split on the 2D:4D measurements to compare individuals with relatively low versus high prenatal testosterone exposure. GEE analyses computed in both groups separately showed that subjects with low prenatal testosterone exposure (high 2D:4D) contributed more to the group after testosterone administration (Wald $\chi^2 = 7.894$, $P = 0.005$), whereas subjects with high prenatal testosterone exposure (low 2D:4D) showed no change (Wald $\chi^2 = 1.791$, $P = 0.181$) (Fig. 1b). A forced-choice test establishing that subjects were unaware of treatment condition also revealed no belief effects on public good contributions (all P values > 0.10). Thus, unlike in the Eisenegger *et al.* study, folk beliefs about testosterone did not mediate behaviour in the PGG¹, which is not unexpected because unfair ultimatum game offers (and not PGG non-contributions) are antisocial and risky, and fit with mainstream ideas on how testosterone affects behaviour^{9,10}.

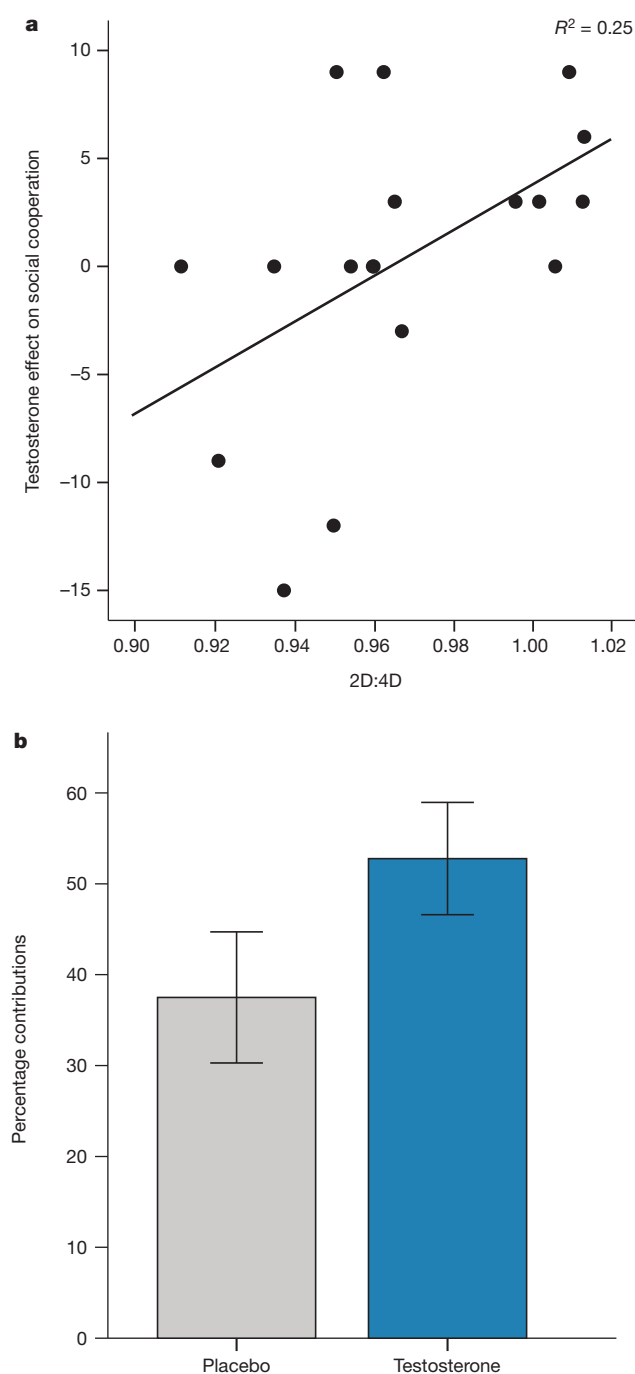


Figure 1 | Prenatal sex hormone priming (2D:4D) moderates effects of testosterone on social cooperation. **a**, Individual 2D:4D measurements plotted against testosterone effect on social cooperation (mean amount of MU contributed in the placebo condition subtracted from mean amount of MU contributed after testosterone administration). Line depicts regression wherein 2D:4D ratio explains 25% of the variance in the overall effect of testosterone on cooperation. **b**, Mean and standard error of the mean of the percentage of overall PGG contribution after testosterone and placebo administration in subjects with relatively low prenatal testosterone (high 2D:4D) on the basis of median split. The data show significantly more overall contribution to the public good after administration of testosterone compared to placebo in high 2D:4D subjects. We thank D. Tromp for her help in collecting these data.

BRIEF COMMUNICATIONS ARISING

The present result supports past research in which we also show effects of testosterone that vary strongly with prenatal testosterone exposure. In that case, high prenatal testosterone exposure (low 2D:4D) boosted the negative effect of testosterone administration on cognitive empathy⁸. Crucially, the 2D:4D ratio apparently is interactively shaped by testosterone and oestradiol *in utero*, and a high 2D:4D ratio points to relatively low prenatal testosterone versus high prenatal oestradiol^{11–13}. Furthermore, many effects of testosterone on social behaviour are thought to arise after metabolism to oestradiol^{2,3}, but this metabolism differs between individuals¹⁴. Hypothetically, the balance between the sex steroids prenatally, marked by the 2D:4D ratio^{11–13}, would be predictive for the rate of metabolism of testosterone into oestradiol. That is, subjects who are prenatally more strongly primed by oestradiol may also metabolize more testosterone into oestradiol¹³, and this could have caused the selective effect in our high 2D:4D group. Further research is necessary to test this hypothesis.

Here we challenge Eisenegger *et al.* by establishing the positive effects of testosterone on social cooperation in which prenatal sex hormone priming, approximated by determining 2D:4D ratios, conveys important individual variability. These data have strong implications for past and future hormone research.

METHODS SUMMARY

For the PGG, all three players receive 3 MU per round and can contribute all or nothing to the public good. Only when at least two players contribute does each player receive an extra 6 MU irrespective of whether they made a contribution; thus non-contributors can profit most (9 MU). To create three-person groups, confederates were used to ensure there were three players involved each time, and their decisions were randomized. After the experiment we checked that no suspicions were raised about this procedure (six non-believers were excluded from analyses). For 2D:4D measurements, subjects' right hands were scanned, and 2D:4D ratios were computed twice by an experienced rater¹⁵ (correlation between measurements, $P < 0.0001$).

Jack van Honk^{1,2}, Estrella R. Montoya¹, Peter A. Bos^{1,2}, Mark van Vugt³ & David Terburg^{1,2}

¹Department of Experimental Psychology, Utrecht University, 3584CS Utrecht, Netherlands.
email: j.vanhonk@uu.nl

²Department of Psychiatry, University of Cape Town, 7925 Cape Town, South Africa.

³Department of Social and Organizational Psychology, VU University, 1081 BT Amsterdam, Netherlands.

Received 19 October 2011; accepted 13 March 2012.

1. Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M. & Fehr, E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* **463**, 356–359 (2010).
2. Bos, P. A., Panksepp, J., Bluthé, R. M. & van Honk, J. Acute effects of steroid hormones and neuropeptides on human social-emotional behavior: a review of single administration studies. *Front. Neuroendocrinol.* **33**, 17–35 (2012).
3. Eisenegger, C., Haushofer, J. & Fehr, E. The role of testosterone in social interaction. *Trends Cogn. Sci.* **15**, 263–271 (2011).
4. Van Vugt, M., De Cremer, D. & Janssen, D. P. Gender differences in cooperation and competition: the male-warrior hypothesis. *Psychol. Sci.* **18**, 19–23 (2007).
5. Tuiten, A. *et al.* Time course of effects of testosterone administration on sexual arousal in women. *Arch. Gen. Psychiatry* **57**, 149–153 (2000).
6. Breedlove, S. M. Organizational hypothesis: instances of the fingerpost. *Endocrinology* **151**, 4116–4122 (2010).
7. Hönekopp, J., Bartholdt, L., Beier, L. & Liebert, A. Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: new data and a meta-analytic review. *Psychoneuroendocrinology* **32**, 313–321 (2007).
8. van Honk, J. *et al.* Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proc. Natl Acad. Sci. USA* **108**, 3448–3452 (2011).
9. Dabbs, J. M. & Dabbs, M. G. *Heroes, Rogues and Lovers: On Testosterone and Behavior* (McGraw-Hill, 2000).
10. van Honk, J. *et al.* Testosterone shifts the balance between sensitivity for punishment and reward in healthy young women. *Psychoneuroendocrinology* **29**, 937–943 (2004).
11. Lutchmaya, S., Baron-Cohen, S., Raggatt, P., Knickmeyer, R. & Manning, J. T. 2nd to 4th digit ratios, fetal testosterone and estradiol. *Early Hum. Dev.* **77**, 23–28 (2004).
12. Manning, J. T., Scutt, D., Wilson, J. & Lewis-Jones, D. I. The ratio of 2nd to 4th digit length: a predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen. *Hum. Reprod.* **13**, 3000–3004 (1998).
13. Zheng, Z. & Cohn, M. J. Developmental basis of sexually dimorphic digit ratios. *Proc. Natl Acad. Sci. USA* **108**, 16289–16294 (2011).
14. Sarachana, T., Xu, M., Wu, R. C. & Hu, V. W. Sex hormones in autism: androgens and estrogens differentially and reciprocally regulate RORA, a novel candidate gene for autism. *PLoS ONE* **6**, e17116 (2011).
15. Millet, K. & Dewitte, S. Second to fourth digit ratio and cooperative behavior. *Biol. Psychol.* **71**, 111–115 (2006).

Author Contributions All authors were involved in the design of the research; D.T. and E.R.M. analysed data; and all authors were involved in the writing of the paper.

Competing Financial Interests Declared none.

doi:10.1038/nature11136

Eisenegger *et al.* reply

REPLYING TO J. van Honk, E. R. Montoya, P. A. Bos, M. van Vugt & D. Terburg *Nature* **485**, <http://dx.doi.org/10.1038/nature11136> (2012)

van Honk and colleagues have taken our findings on the role of testosterone in ultimatum game bargaining¹ a step forward by showing that the hormone has important prosocial effects beyond the ultimatum game by increasing cooperation in the public goods game (PGG)². In contrast to the ultimatum game, participants in the PGG decide simultaneously about their cooperation levels and are not confronted with a rejection threat from other participants, suggesting a much more universal effect of testosterone on prosociality than revealed by our study¹. As the PGG captures a large class of evolutionarily and contemporaneously important situations, their findings are of great interest, and considerably extend our knowledge about the causal effect of testosterone on social behaviour. In addition, their results raise intriguing questions regarding the motivational and biological mechanisms through which testosterone increased cooperation levels, suggesting that the study will trigger further important experiments.

One possibility is that testosterone directly affects people's social preferences, that is, it directly renders their motives more prosocial—an explanation that seems to be favoured by van Honk and colleagues². A second possibility is that the hormone influences subjects' beliefs about the other players' cooperation levels. For example, a subject who received testosterone may believe that others are more cooperative while their own motives remain unchanged. The rules of the PGG used in van Honk *et al.* are such that it is in the third player's financial interest to contribute to the public good if that subject believes that precisely one of the two other players contributes her endowment as well. Thus it is possible that the positive effect of testosterone on cooperativeness is due to the hormone's effect on beliefs about other participants' cooperation level. Of course, the two mechanisms described above are not mutually exclusive; testosterone may have causally affected subjects' motives and beliefs about others' cooperation. So far, no study has examined testosterone's causal effects

BRIEF COMMUNICATIONS ARISING

on beliefs about others' social preferences or others' prosocial behaviour—suggesting this as a topic for future studies.

Interestingly, the results of van Honk *et al.* are also compatible with the hypothesis that testosterone increases concerns for one's status^{3,4}. One study⁵ has shown that participants in a PGG confer higher status to cooperative group members than to non-cooperative ones. Another study⁶ reported that third-party observers who did not benefit in any way from high contributions also rated cooperative individuals higher on status. Thus, if subjects have internalized the link between cooperation in the PGG and their status in the group, the increased cooperation levels observed by van Honk *et al.* could have been caused by an increased concern about one's status.

The study by van Honk and colleagues also raises intriguing questions regarding the biological mechanisms underlying the hormone's effect on social behaviour. As testosterone is partially converted to oestradiol by the enzyme aromatase, the results obtained by van Honk *et al.* suggest that testosterone's effect on cooperation in the PGG in subjects with a high digit ratio might be mediated by oestradiol. This idea is based on a recent study which shows that the digit ratio is related to an individual's relative amount of androgen to oestrogen activity during early fetal development, and that oestradiol heightens the digit ratio⁷. This suggests that effects of testosterone on social behaviour may be mediated by an individual's relative androgen versus oestrogen system activity or sensitivity, for which the digit ratio might serve as an important biomarker^{7,8}. Thus, future studies should test whether a high digit ratio directly predicts social cooperation after oestradiol administration, as this might provide further insights into the biological mechanism behind the behavioural impact of sex hormones.

The study by van Honk *et al.* considerably broadens the causal role of testosterone on important prosocial behaviours while simultaneously suggesting a more specific biological underpinning based on the idea that the digit ratio is a biomarker for individuals' androgen

versus oestrogen system activity or sensitivity. In addition, the study raises important questions regarding the mechanisms underlying the relationship between testosterone administration and prosocial behaviour. In particular, studies that are able to distinguish between the three hypotheses—that testosterone affects social preferences, beliefs, or status seeking—are likely to move the field forward further.

C. Eisenegger¹, M. Naef^{1,2}, R. Snozzi¹, M. Heinrichs³ & E. Fehr¹

¹Department of Economics, Laboratory for Social and Neural Systems Research, University of Zurich, 8006 Zurich, Switzerland.

email: christoph.eisenegger@gmail.com

²Department of Economics, Royal Holloway, University of London, Egham TW20 0EX, UK.

³Department of Psychology, Laboratory for Biological and Personality Psychology, University of Freiburg, 79104 Freiburg, Germany.

1. Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M. & Fehr, E. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* **463**, 356–359 (2010).
2. van Honk, J., Montoya, E. R., Bos, P. A., van Vugt, M. & Terburg, D. New evidence on testosterone and cooperation. *Nature* **485**, <http://dx.doi.org/10.1038/nature11136> (2012).
3. Mazur, A. & Booth, A. Testosterone and dominance in men. *Behav. Brain Sci.* **21**, 353–397 (1998).
4. Eisenegger, C., Haushofer, J. & Fehr, E. The role of testosterone in social interaction. *Trends Cogn. Sci.* **15**, 263–271 (2011).
5. Hardy, C. L. & Van Vugt, M. Nice guys finish first: the competitive altruism hypothesis. *Pers. Soc. Psychol. Bull.* **32**, 1402–1413 (2006).
6. Willer, R. Groups reward individual sacrifice: the status solution to the collective action problem. *Am. Sociol. Rev.* **74**, 23–43 (2009).
7. Zheng, Z. & Cohn, M. J. Developmental basis of sexually dimorphic digit ratios. *Proc. Natl Acad. Sci. USA* **108**, 16289–16294 (2011).
8. van Honk, J. *et al.* Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. *Proc. Natl Acad. Sci. USA* **108**, 3448–3452 (2011).

doi:10.1038/nature11137

FORUM: Climate change

Flowering in the greenhouse

Predicting plant responses to increasing temperatures is integral to assessing the global impact of climate change. But the authors of a comparative study assert that warming experiments may not accurately reflect observational data. Climate and ecosystem scientists discuss how impact prediction should proceed. [SEE LETTER P.494](#)

THE PAPER IN BRIEF

- Seasonal plant flowering and leafing have occurred earlier in recent decades, almost certainly as a result of higher average temperatures.
- Artificial warming experiments are designed to replicate changed climatic conditions in a controlled environment, allowing direct assessment of

the effects on plants.

- Wolkovich *et al.*¹ (page 494) compare existing data on the responses of 1,634 plant species, across four continents, to observed and experimental warming.
- They find that observed plant flowering and leafing dates have occurred significantly earlier than the changes calculated from artificial warming experiments.

Consider all data

THIS RUTISHAUSER & RETO STÖCKLI

Phenology, the science of seasonally recurring life-cycle stages in plants and animals, has become an important aspect of climate science. Anthropogenic climate change is leading to anomalous conditions, including warming trends that strongly correlate with changing plant flowering and leafing dates. In turn, these unusual phenological manifestations may lead to anomalous bird migration and to plant pollination patterns that could desynchronize entire food chains and have socio-economic impacts that remain poorly defined. Observational studies are needed to document and understand these changes, and such studies are essential to constrain models used to estimate the effects of projected future warming on Earth's carbon and water cycles.

One open question is whether temperature sensitivities derived from past phenological observations will remain valid in a future, warmer climate. To answer this question, biologists typically rely on plant temperature sensitivities (defined as the number of days' change in flowering or leafing date per degree of temperature change) determined from controlled, plot-scale field experiments in which vegetation is warmed artificially. Wolkovich and her interdisciplinary team¹ compared the sensitivities ascertained from long-term observational records with results from

warming experiments. They compiled data on temperature sensitivities of 1,634 plant species across multiple latitudes and temporal scales, which dated as far back as the nineteenth century. They found that flowering and leafing sensitivities obtained from warming experiments systematically underpredict those from long-term observational records by a factor of 8.5 and 4.0, respectively. Their analysis shows that these mismatches do not depend on latitude and that there is little evidence that experimental design influenced the degree of underprediction.

It is not surprising that plant responses to warming-only field experiments are lower than to real-world variance occurring over many decades. Temperature changes are not independent of other environmental changes — warm spring seasons, for example, are often accompanied by

more sunshine, drier soils and shortened snow duration. All of these factors influence plant growth and phase-change events, such as budding and flowering². Moreover, temperature sensitivities are neither constant in time nor fixed within a species, and can vary by between 3 and 6 days per degree when data recorded over centuries are considered³.

The difficulty in replicating these factors experimentally means that Wolkovich and colleagues' call for ecosystem-model re-evaluation must be taken seriously. Current

vegetation models predict either a delayed or lengthened plant growing season, either of which will have adverse effects on the terrestrial carbon uptake⁴. But recent studies have revealed substantial deficiencies in these models. For example, exceptionally warm winter temperatures can disrupt a plant's requirements for chilling, and thus, apparently paradoxically, delay leafing and flowering⁵. Phenology is much more than a linear correlate to annual mean air temperatures — for a model to make realistic phenological predictions, it must combine the effects of, at least, temperature, light and moisture⁶.

The real and potential impacts of phenological change on natural ecosystems attract attention from scientists and the public alike, as, for example, in the interview⁷, 'What's the impact of early blooms?', which was broadcast in March on the US National Public Radio programme *All Things Considered*, and the burgeoning number of 'citizen science' projects, in which individuals provide researchers with observational field data (see, for example, refs 8–10). The prediction discrepancies revealed by Wolkovich and colleagues can be resolved only by further analyses of long-term phenological observations in combination with climate and ecosystem data, backed by model-based and experimental results. The key message of this study is that we must consider, and obtain, more observational data to better understand the processes of the global 'warming experiment' that is ongoing on our planet — one that is taking place daily, without control plots and on the most relevant scale. We must put all these data to use.

This Rutishauser is at the Oeschger Centre for Climate Change Research, the Institute of Geography, University of Bern, and the Commission for Phenology and Seasonality, Swiss Academies of Sciences SCNAT, CH-3012 Bern, Switzerland.

Reto Stöckli is at the Federal Office of Meteorology and Climatology MeteoSwiss, CH-8044 Zurich, Switzerland.
e-mails: rutis@giub.unibe.ch;
reto.stoeckli@meteoswiss.ch

Insight from integration

JOHN HARTE & LARA KUEPPERS

Anthropogenic climate change, a massive uncontrolled planetary disruption, is under way, and a predictive understanding of how it will affect ecosystems and how these ecosystem responses will feed back to the climate is badly needed. Meta-analyses, such as that presented by Wolkovich *et al.*¹, have the potential to augment current ecological understanding. But when such analyses fail to explicitly address methodological differences between data sources that are crucial for interpreting results, their conclusions can be misleading. Here, we urge caution in the use of meta-analyses to assess different approaches to predicting ecosystem responses to climate change, and we describe how integrating multiple investigative methods can better advance understanding.

To predict ecosystem responses, scientists can observe correlations between climate and ecosystem properties, and can conduct controlled experimental manipulations of climate. However, both methods have limitations as well as strengths — it is difficult, for example, to identify causal mechanisms from correlative observations, and controlled experiments on small spatial scales cannot reflect all aspects of larger environments. Predictions of ecosystem responses should strive to make use of information from differing approaches in a way that considers the strengths and drawbacks of each.

Simply comparing results from many different experimental and correlative studies, as Wolkovich and colleagues attempted, can obscure nuances of methodology and analysis that are essential for making accurate predictions. Alternatively, integrating observational and experimental methods within a single study¹¹ can simultaneously provide insight into issues that are obscured by meta-analyses and the mechanisms driving ecosystem responses. For example, we have studied the sensitivity of subalpine meadow plants to climate using two types of experimental climate change (year-round heating and snow removal) and two types of natural climate change (interannual variability and change along an elevation gradient). We found¹² that the timing of flowering, and many other ecological responses, are remarkably similarly sensitive to each of these four sources of climate change, provided that snow-melt date is used as an explanatory variable, rather than mean annual temperature, as used by Wolkovich and colleagues.

The type of experimental method used for such studies does matter, though. Had we used passive heating methods, such as open-top chambers (Fig. 1) deployed only during the growing season, we might have increased



Figure 1 | Early bloomers. Artificial warming experiments, such as that shown here in the White Mountains of California, in which plants growing in open-top warming chambers are compared with adjacent plants growing in normal conditions, can be used to study how plant flowering and leafing times alter in different temperature conditions, and thereby help to predict how plants will respond to climate change.

the mean annual temperature, but our experimental heating would not have advanced snow melt and we would not have observed a strong phenological response. Thus, meta-analyses should carefully consider both the appropriate explanatory variables and methodological differences across experimental studies. Explanatory variables are likely to differ from one type of habitat to another, and from one type of

“Explanatory variables are likely to differ from one type of habitat to another.”

response variable to another. In lowland tropical regions, where snow is not a factor, there is a great need for integrative methods to determine which variables drive plant and animal, as well as biogeochemical, responses to climate change.

Our observation that phenology responds to changing snow-melt date consistently across different observational and experimental methods might suggest that researchers could dispense with expensive warming experiments and rely solely on cheaper, easier correlative studies. But that would be a mistake. Diverse ecosystem responses play out over a range of timescales and require persistent multi-year manipulations to uncover causal mechanisms¹³. For example, the change in soil carbon levels that occurs over time in response to persistent shifts in plant-species composition is not easily captured by interannual climate variability, but we recorded it in our controlled 21-year warming experiment¹⁴. Furthermore, predicting climate impacts on species turnover, selection and adaptation may require integration of yet other types of observation and experiment, such as the use of data from the fossil record and reciprocal transplant experiments across climate gradients.

These examples demonstrate how controlled warming experiments, particularly those that proceed for longer than the typical research-funding cycle of 3–5 years, can provide a means of understanding both patterns in ecosystem responses and the multiple mechanisms that govern these changes. Such experiments will allow identification and manipulation of explanatory variables and, if they are performed in combination with observational studies that take advantage of climate variability in space and time, will guide our quest to predict the future of ecosystems under global change. ■

John Harte is at the University of California, Berkeley, Berkeley, California 94720, USA.

Lara Kueppers is at the University of California, Merced, Merced, California 95343, USA.

e-mails: jharte@berkeley.edu; lkueppers@ucmerced.edu

- Wolkovich, E. M. *et al.* *Nature* **485**, 494–497 (2012).
- Körner, C. in *Plant Growth and Climate Change* (eds Morison, J. I. L. & Morecroft, M. D.) 48–69 (Blackwell, 2006).
- Rutishauser, T., Luterbacher, J., Defila, C., Frank, D. & Wanner, H. *Geophys. Res. Lett.* **35**, L05703 (2008).
- Richardson, A. D. *et al.* *Glob. Change Biol.* **18**, 566–584 (2012).
- Vitasse, Y. C. *et al.* *Agric. Forest Meteorol.* **151**, 969–980 (2011).
- Stöckli, R., Rutishauser, T., Baker, I., Liniger, M. A. & Denning, A. S. *J. Geophys. Res.* **116**, G03020 (2011).
- www.npr.org/2012/03/15/148698272/whats-the-impact-of-early-blooms
- www.phaeno.ethz.ch/globe
- www.blommar.nu
- naturescalendar.org.uk
- Dunne, J. A., Saleska, S. R., Fischer, M. L. & Harte, J. *Ecology* **85**, 904–916 (2004).
- Dunne, J. A., Harte, J. & Taylor, K. J. *Ecol. Monogr.* **73**, 69–86 (2002).
- Smith, M. D., Knapp, A. K. & Collins, S. L. *Ecology* **90**, 3279–3289 (2009).
- Saleska, S. R. *et al.* *Glob. Biogeochem. Cycles* **16**, 1055 (2002).

Molecules meet materials

An inorganic semiconductor can take the place of the liquid electrolyte typically used in dye-sensitized solar cells. This achievement points the way to making these devices more easily manufactured and more stable. [SEE LETTER P.486](#)

THOMAS E. MALLOUK

Dye-sensitized solar cells are inexpensive devices that convert sunlight into electricity with 10–12% efficiency. The design of the cell is simple. It consists of a thick film of porous titanium dioxide covered with a monolayer of light-absorbing dye molecules and sandwiched between two plates of conducting glass. This simplicity makes dye cells an attractive alternative to conventional solid-state cells. However, the liquid electrolyte in which the dye molecules are immersed complicates the manufacture of the cells and reduces their stability. This problem was recognized early on, when these dye cells were first introduced¹ in 1991, and a solid-state version² followed in 1998. However, high efficiencies for solid-state dye cells have not been achieved. On page 486 of this issue, Chung *et al.*³ show that an inorganic semiconductor can be used in place of the liquid electrolyte to make efficient solid-state dye cells.

For a dye-sensitized solar cell to work, the monolayer of light-absorbing molecules must be positioned at the interface between

the titanium dioxide film and a conductor of 'holes' (conceptual particles formed by the absence of electrons). Complete light absorption requires a rough titanium dioxide film that has a large surface area; with a planar film the monolayer can absorb only a little of the incident light. Unlike liquid electrolytes, the molecular glasses, conducting polymers and semiconductors that can permeate the film's nanopore network for conduction tend to be poor hole conductors. In these conductors, many of the electrons and holes that are created by light absorption annihilate each other before they can escape from the pore network to the cell's cathode and anode, respectively, to allow electric current to be generated.

Chung and colleagues' inorganic semiconductor, a fluorine-doped caesium tin iodide ($\text{CsSnI}_{2.95}\text{F}_{0.05}$), is one of a family of tin- and lead-halide perovskite compounds that has been studied for many years for use in solution-processable semiconductors⁴. When dissolved in an organic solvent such as dimethylformamide, it crystallizes to fill the pores of the titanium dioxide film and provides a continuous pathway for hole conduction. The high

hole mobility of the material results in cells whose electric current and voltage resemble those of a good liquid-based dye cell. The performance of the authors' cells is encouraging: not only is their internal electrical resistance low, which is to be expected for cells that have good hole conductors, but they also achieve high current and voltage. The latter parameters imply that electrons from the titanium dioxide transfer slowly across the titanium dioxide–dye–semiconductor interface, and thus only a small proportion of the electrons generated by light absorption are lost to the electron–hole annihilation process (Fig. 1).

In liquid-based dye cells, slow electron–hole annihilation must be deliberately engineered by using redox molecules in the liquid electrolyte (such as iodide or cobalt complexes) that accept electrons from titanium dioxide at slow rates, or by blocking the access of these molecules to the surface of the titanium dioxide film using bulky organic monolayers⁵. In early versions of solid-state dye cells, the annihilation process between electrons in the titanium dioxide film and holes in the glassy organic hole conductors was easy to understand. According to the classical Marcus theory of electron-transfer reactions, if molecules do not move or distort much during an electron-transfer reaction, then the reaction tends to be fast, and this was the case with the glassy organic hole conductors. Why the annihilation process is slow in the new solid-state cell is unknown.

The operating principles of a liquid-based dye cell are fundamentally different from those of a solid-state solar cell, in which light is absorbed in a semiconductor structure known as a p–n junction. In a p–n-junction cell, such as a conventional silicon solar cell, the junction's built-in electric field sweeps charge-carrying electrons and holes in opposite directions. This is a good thing, because the minority charge carriers are born in 'enemy territory', where they must escape from the majority carriers before they annihilate each other. In a liquid-based dye cell, all the carriers are generated at the titanium dioxide–electrolyte interface, and there is no appreciable electric field because of charge screening by the electrolyte⁶. Electrons injected into the titanium dioxide by the dye move by diffusion, and electron–hole annihilation can happen whenever the electrons wander back to the interface.

By contrast, in Chung and colleagues' cell, there is a new kind of interface, namely two solid semiconductors (the titanium dioxide and the fluorine-doped caesium tin iodide), separated by a single layer of dye molecules. Such an interface might sustain an electric field in the fluorine-doped caesium tin iodide that would help electrons and holes to escape from one other. Indeed, the cell does exhibit some light-induced current without any dye being present, indicating that electrons and holes, created in this case by light absorption

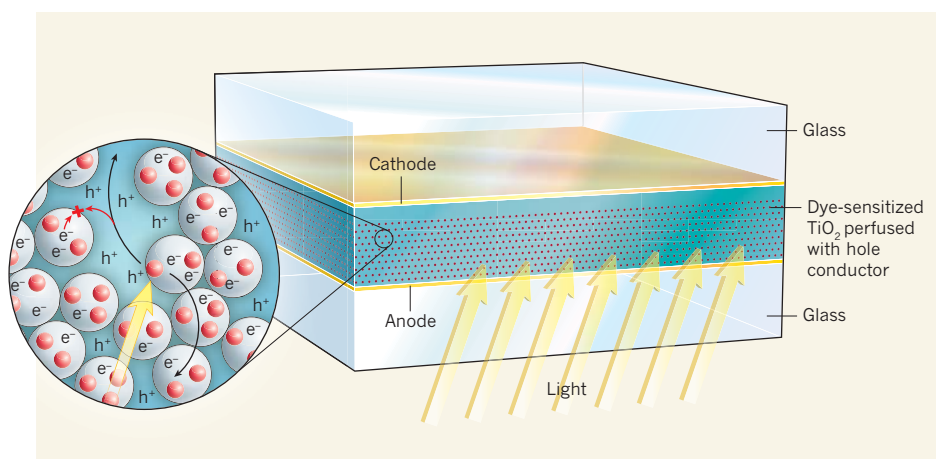


Figure 1 | A solid-state dye cell. Chung and colleagues' dye-sensitized solar cell³ consists of a porous film of titanium dioxide (TiO_2) particles (white; inset) coated with light-sensitizing dye molecules (red) and filled with a solid-state material that conducts 'holes'. This hole conductor is sandwiched between two plates of glass, one attached to the cell's cathode and the other to the anode. When light is shone on the cell, electrons (e^-) introduced into the titanium dioxide by the excited dye molecules diffuse to the anode and holes (h^+) in the hole conductor travel to the cathode. For the cell to convert the light efficiently into electric current, the rate at which electrons and holes encounter and annihilate one another (red arrows) must be slow relative to the rate of hole transport. The TiO_2 film is about 10 micrometres thick.

in caesium tin iodide, do separate. With the dye, one has the best of both worlds: all the electrons and holes are generated at the interface and are transferred in opposite directions.

Tuning the energy levels of dyes and redox molecules has led⁵ to efficient liquid-based dye cells, with voltages approaching 1 volt. But further progress has been constrained by the need for redox molecules that are slow at accepting electrons from the titanium dioxide and also well matched to the redox level of the dye — often conflicting requirements. For example, the iodide–tri-iodide redox couple accepts electrons slowly from the titanium dioxide, but does so in a process that involves making and breaking the I–I bond of tri-iodide and costs energy. This translates into a lower voltage for iodide-based dye cells than they would have if bond making and breaking could be avoided. If electron–hole annihilation can be circumvented by a different mechanism in the

solid-state titanium dioxide–dye–semiconductor cell, other than by making and breaking chemical bonds, then there should be broader latitude for tuning the energetics of the cells and increasing their voltage. The fact that a range of solid solutions can be made in the perovskite family of solution-processable semiconductors suggests that elaborating the chemistry of this class of hole conductor may be a fruitful avenue for dye–cell research. ■

Thomas E. Mallouk is in the Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, USA. e-mail: tem5@psu.edu

1. O'Regan, B. & Grätzel, M. *Nature* **353**, 737–740 (1991).
2. Bach, U. *et al.* *Nature* **395**, 583–585 (1998).
3. Chung, I. *et al.* *Nature* **485**, 486–489 (2012).
4. Mitzi, D. B. *Prog. Inorg. Chem.* **48**, 1–121 (1999).
5. Yella, A. *et al.* *Science* **334**, 629–634 (2011).
6. Gregg, B. A. *J. Phys. Chem. B* **107**, 4688–4698 (2003).

ALZHEIMER'S DISEASE

A breach in the blood–brain barrier

Alterations in brain blood vessels in mice precede the neural dysfunction associated with Alzheimer's disease. The finding highlights potential targets for drug development. SEE LETTER P.512

PETER CARMELIET & BART DE STROOPER

The loss of function and death of neurons that occur in patients with Alzheimer's disease are traditionally considered^{1,2} to result from the accumulation in the brain of a peptide known as amyloid- β . Little attention has been paid to the possible role of the vascular lesions that are also observed in such patients^{3,4}, because these lesions are generally viewed as a consequence rather than a cause of the disorder. However, it has long been known that *APOE4*, a variant of the gene that encodes apolipoprotein E (APOE), is strongly associated with Alzheimer's disease⁵. The protein is involved in the binding, transport and metabolism of lipids, and in the clearance of amyloid- β from the brain⁶.

On page 512 of this issue, Bell *et al.*⁷ report that, in mice, the protein variant APOE4 unleashes a destructive inflammatory response in pericytes — cells that wrap around blood vessels and seal off the blood–brain barrier (BBB). The resulting breach in the barrier leads to leakage into the brain of blood-borne molecules that are toxic to neurons and cause neurodegenerative changes. Furthermore, the researchers provide proof of principle that pharmacological targeting of inflamed pericytes

can heal the BBB and repair neural damage.

Some factors that increase the risk of vascular disorders also seem to increase the risk of Alzheimer's disease^{3,4}, although the relevance of these findings is controversial. Work in animal models^{3,8} nonetheless supports the hypothesis that certain alterations in pericytes that lead to disturbed function of the BBB might indeed be an early and perhaps causal event in Alzheimer's disease. To explore the role of APOE in these processes, Bell *et al.* replaced the native *ApoE* gene in mice with one of several variants of human APOE (*APOE2*, *APOE3* or *APOE4*). They found that the animals carrying *APOE4*, but not those expressing *APOE2* or *APOE3* (which are not associated with Alzheimer's disease), had a disrupted BBB, fewer blood vessels and reduced blood flow in some parts of the brain.

The authors hypothesized that a pro-inflammatory cell-signalling molecule called cyclophilin A (CypA) could mediate BBB disruption in the APOE4-expressing mice, because in mouse models of vascular disorders⁹ CypA has deleterious effects on endothelial cells — which form the walls of blood vessels — and on pericytes. Indeed, Bell *et al.* report that, in APOE4-expressing animals and in mice lacking any APOE proteins,

CypA initiates a pro-inflammatory pathway driven by the transcription factor NF- κ B (see Fig. 4e of the paper⁷). This, in turn, results in elevated levels of MMP9, a proteinase enzyme that breaks down the tight junctions between adjacent endothelial cells as well as the protein components of the basement membrane that surrounds blood vessels. Neurotoxic molecules leak out of the vessels, damaging neurons and affecting their synaptic connections.

Bell *et al.* find that APOE3 — which does not trigger BBB dysfunction — binds to a receptor protein (LRP1) on the surface of pericytes and, by doing so, prevents the inflammatory response. By contrast, in mice expressing APOE4 or lacking any APOE proteins, this interaction does not occur, and the resulting inflammatory response breaks down the BBB. The authors' findings suggest that APOE4 cannot bind to LRP1 as efficiently as does APOE3, and so fails to 'cool down' the brain pericytes. Of note, the researchers show that the APOE4-triggered changes in the BBB precede neurological deterioration in the mice studied. Moreover, chemical inhibitors of this pro-inflammatory pathway healed the breached BBB and ameliorated the neuronal changes in the animals, opening up the possibility of improved treatment of Alzheimer's disease.

The study raises several interesting questions. For instance, how important are pericytes in preventing disease initiation or progression? Is it possible that subtle neuronal damage occurred before the vascular changes began but was not detected by the authors? The pathway described by Bell *et al.* could work in some patients with Alzheimer's disease who produce APOE4 and not APOE3. However, as it is the lack of APOE3 rather than the presence of APOE4 that triggers the inflammatory response, the role of this pathway in the many patients with Alzheimer's disease who carry one copy of *APOE4* and one of *APOE3* is unclear. Moreover, why are pericytes under continuous threat of becoming 'inflamed' by the CypA–NF- κ B pathway, and what are the signals that activate this pathway?

Apart from these general considerations, the current work⁷ did not explore whether the vascular and neurological changes observed in the brains of APOE4-expressing mice were accompanied by accumulation of amyloid- β . Previous work⁶ has shown that this peptide accumulates in the brains of patients carrying *APOE4* and in *APOE4*-expressing mice, but the link between amyloid- β accumulation and vascular alterations is unclear.

How relevant are these findings for the treatment of Alzheimer's disease? Bell and colleagues' results suggest that several protein components of the CypA–NF- κ B pathway may be potential targets for the development of drugs that could normalize both the BBB defects and the neural damage. Moreover, pharmacological activation of the retinoid X receptor — a protein that stimulates APOE expression — reversed

in caesium tin iodide, do separate. With the dye, one has the best of both worlds: all the electrons and holes are generated at the interface and are transferred in opposite directions.

Tuning the energy levels of dyes and redox molecules has led⁵ to efficient liquid-based dye cells, with voltages approaching 1 volt. But further progress has been constrained by the need for redox molecules that are slow at accepting electrons from the titanium dioxide and also well matched to the redox level of the dye — often conflicting requirements. For example, the iodide–tri-iodide redox couple accepts electrons slowly from the titanium dioxide, but does so in a process that involves making and breaking the I–I bond of tri-iodide and costs energy. This translates into a lower voltage for iodide-based dye cells than they would have if bond making and breaking could be avoided. If electron–hole annihilation can be circumvented by a different mechanism in the

solid-state titanium dioxide–dye–semiconductor cell, other than by making and breaking chemical bonds, then there should be broader latitude for tuning the energetics of the cells and increasing their voltage. The fact that a range of solid solutions can be made in the perovskite family of solution-processable semiconductors suggests that elaborating the chemistry of this class of hole conductor may be a fruitful avenue for dye-cell research. ■

Thomas E. Mallouk is in the Department of Chemistry, Pennsylvania State University, University Park, Pennsylvania 16802, USA. e-mail: tem5@psu.edu

1. O'Regan, B. & Grätzel, M. *Nature* **353**, 737–740 (1991).
2. Bach, U. *et al.* *Nature* **395**, 583–585 (1998).
3. Chung, I. *et al.* *Nature* **485**, 486–489 (2012).
4. Mitzi, D. B. *Prog. Inorg. Chem.* **48**, 1–121 (1999).
5. Yella, A. *et al.* *Science* **334**, 629–634 (2011).
6. Gregg, B. A. *J. Phys. Chem. B* **107**, 4688–4698 (2003).

ALZHEIMER'S DISEASE

A breach in the blood–brain barrier

Alterations in brain blood vessels in mice precede the neural dysfunction associated with Alzheimer's disease. The finding highlights potential targets for drug development. SEE LETTER P.512

PETER CARMELIET & BART DE STROOPER

The loss of function and death of neurons that occur in patients with Alzheimer's disease are traditionally considered^{1,2} to result from the accumulation in the brain of a peptide known as amyloid- β . Little attention has been paid to the possible role of the vascular lesions that are also observed in such patients^{3,4}, because these lesions are generally viewed as a consequence rather than a cause of the disorder. However, it has long been known that *APOE4*, a variant of the gene that encodes apolipoprotein E (APOE), is strongly associated with Alzheimer's disease⁵. The protein is involved in the binding, transport and metabolism of lipids, and in the clearance of amyloid- β from the brain⁶.

On page 512 of this issue, Bell *et al.*⁷ report that, in mice, the protein variant APOE4 unleashes a destructive inflammatory response in pericytes — cells that wrap around blood vessels and seal off the blood–brain barrier (BBB). The resulting breach in the barrier leads to leakage into the brain of blood-borne molecules that are toxic to neurons and cause neurodegenerative changes. Furthermore, the researchers provide proof of principle that pharmacological targeting of inflamed pericytes

can heal the BBB and repair neural damage.

Some factors that increase the risk of vascular disorders also seem to increase the risk of Alzheimer's disease^{3,4}, although the relevance of these findings is controversial. Work in animal models^{3,8} nonetheless supports the hypothesis that certain alterations in pericytes that lead to disturbed function of the BBB might indeed be an early and perhaps causal event in Alzheimer's disease. To explore the role of APOE in these processes, Bell *et al.* replaced the native *ApoE* gene in mice with one of several variants of human APOE (*APOE2*, *APOE3* or *APOE4*). They found that the animals carrying *APOE4*, but not those expressing *APOE2* or *APOE3* (which are not associated with Alzheimer's disease), had a disrupted BBB, fewer blood vessels and reduced blood flow in some parts of the brain.

The authors hypothesized that a pro-inflammatory cell-signalling molecule called cyclophilin A (CypA) could mediate BBB disruption in the APOE4-expressing mice, because in mouse models of vascular disorders⁹ CypA has deleterious effects on endothelial cells — which form the walls of blood vessels — and on pericytes. Indeed, Bell *et al.* report that, in APOE4-expressing animals and in mice lacking any APOE proteins,

CypA initiates a pro-inflammatory pathway driven by the transcription factor NF- κ B (see Fig. 4e of the paper⁷). This, in turn, results in elevated levels of MMP9, a proteinase enzyme that breaks down the tight junctions between adjacent endothelial cells as well as the protein components of the basement membrane that surrounds blood vessels. Neurotoxic molecules leak out of the vessels, damaging neurons and affecting their synaptic connections.

Bell *et al.* find that APOE3 — which does not trigger BBB dysfunction — binds to a receptor protein (LRP1) on the surface of pericytes and, by doing so, prevents the inflammatory response. By contrast, in mice expressing APOE4 or lacking any APOE proteins, this interaction does not occur, and the resulting inflammatory response breaks down the BBB. The authors' findings suggest that APOE4 cannot bind to LRP1 as efficiently as does APOE3, and so fails to 'cool down' the brain pericytes. Of note, the researchers show that the APOE4-triggered changes in the BBB precede neurological deterioration in the mice studied. Moreover, chemical inhibitors of this pro-inflammatory pathway healed the breached BBB and ameliorated the neuronal changes in the animals, opening up the possibility of improved treatment of Alzheimer's disease.

The study raises several interesting questions. For instance, how important are pericytes in preventing disease initiation or progression? Is it possible that subtle neuronal damage occurred before the vascular changes began but was not detected by the authors? The pathway described by Bell *et al.* could work in some patients with Alzheimer's disease who produce APOE4 and not APOE3. However, as it is the lack of APOE3 rather than the presence of APOE4 that triggers the inflammatory response, the role of this pathway in the many patients with Alzheimer's disease who carry one copy of *APOE4* and one of *APOE3* is unclear. Moreover, why are pericytes under continuous threat of becoming 'inflamed' by the CypA–NF- κ B pathway, and what are the signals that activate this pathway?

Apart from these general considerations, the current work⁷ did not explore whether the vascular and neurological changes observed in the brains of APOE4-expressing mice were accompanied by accumulation of amyloid- β . Previous work⁶ has shown that this peptide accumulates in the brains of patients carrying *APOE4* and in *APOE4*-expressing mice, but the link between amyloid- β accumulation and vascular alterations is unclear.

How relevant are these findings for the treatment of Alzheimer's disease? Bell and colleagues' results suggest that several protein components of the CypA–NF- κ B pathway may be potential targets for the development of drugs that could normalize both the BBB defects and the neural damage. Moreover, pharmacological activation of the retinoid X receptor — a protein that stimulates APOE expression — reversed

amyloid- β -induced deficits in neural function in a mouse model of Alzheimer's disease, in an APOE-dependent way¹⁰. But the effects on the BBB were not analysed in that study, leaving unanswered the question of whether increased APOE levels, changes in the lipid content of APOE particles, or other mechanisms (for instance, pericyte degeneration, as described by Bell *et al.*) contributed to the beneficial outcome of the pharmacological treatment in that mouse model.

It is tempting to speculate that pericytes fail to guard the BBB in other neurodegenerative diseases, such as Parkinson's disease and amyotrophic lateral sclerosis, in which vascular changes have been documented^{3,4}. So the provocative findings by Bell *et al.* lay down an exciting new path for future research and drug discovery regarding not only the role of APOE4 in Alzheimer's disease, but also the intriguing and fascinating interplay between the vascular and nervous systems in the development of neurodegenerative disorders. ■

Peter Carmeliet is at the *Vesalius Research Center, VIB, University of Leuven*.
Bart De Strooper is at the *VIB Center for the Biology of Disease and the Center for Human Genetics, University of Leuven, Leuven B-3000, Belgium*.
e-mails: peter.carmeliet@vib-kuleuven.be; bart.destrooper@cme.vib-kuleuven.be

1. De Strooper, B. *Physiol. Rev.* **90**, 465–494 (2010).
2. Karran, E., Mercken, M. & De Strooper, B. *Nature Rev. Drug Discov.* **10**, 698–712 (2011).
3. Quaegebeur, A., Lange, C. & Carmeliet, P. *Neuron* **71**, 406–424 (2011).
4. Storkebaum, E., Quaegebeur, A., Vikkula, M. & Carmeliet, P. *Nature Neurosci.* **14**, 1390–1397 (2011).
5. Genin, E. *et al. Mol. Psychiatry* **16**, 903–907 (2011).
6. Castellano, J. M. *et al. Sci. Transl. Med.* **3**, 89ra57 (2011).
7. Bell, R. D. *et al. Nature* **485**, 512–516 (2012).
8. Zlokovic, B. V. *Nature Rev. Neurosci.* **12**, 723–738 (2011).
9. Satoh, K., Shimokawa, H. & Berk, B. C. *Circ. J.* **74**, 2249–2256 (2010).
10. Cramer, P. E. *et al. Science* **335**, 1503–1506 (2012).

Archaean basalts. As Keller and Schoene point out⁴, the latter cannot simply reflect a drop in mantle temperature; instead it suggests a fundamental change in the way that heat moved through, and was lost by, Earth.

Geologists have long known that the Archaean–Proterozoic transition marks one of the most momentous and consequential times in Earth's history⁶ — it immediately followed the most rapid increase in Earth's crustal volume. Petrologists and geochemists have also long noted a difference in the composition of Archaean and post-Archaean continental crust that seems to be accompanied by changes in the thickness of the crust and of the lithosphere (Earth's rigid, outermost layer). A host of ideas has been suggested to explain these changes, ranging from the beginning of modern plate tectonics to an alteration in the character of subduction. One issue with Keller and Schoene's result is that there is a paucity of data from the earliest Proterozoic, but it is hard to see how that alone could explain the discontinuity. Indeed, a lull in volcanic activity in the early Proterozoic, which would explain the lack of data, has previously been noted⁷; the lull itself represents yet another puzzling aspect of the Archaean–Proterozoic transition.

Intriguingly, the Archaean–Proterozoic transition was also a time of pronounced changes at Earth's surface: between 2.4 billion and 2.2 billion years ago, the oxygen content of the atmosphere rose from a few parts per million to about 2% by volume⁸; for comparison, the modern volume fraction is around 20%. Earth also experienced its first major glaciation⁹ — one that was possibly much more severe than the Pleistocene glaciations that have occurred during the past 2 million years. Studies have shown that the oxygenation and the first glaciation occurred at almost the same time^{10,11}. Atmospheric composition exerts a strong influence on climate through the greenhouse effect, so it is not surprising that these events seem to be related. For example, the increase in oxygen might have led to a drop in levels of methane (a potent greenhouse gas), causing the greenhouse effect to crash, thereby tipping Earth into an 'ice-house' state¹². However, a consensus has yet to emerge on exactly what happened.

The fact that such marked changes seem to have occurred almost simultaneously in Earth's interior and at its surface raises the question of whether those changes are linked. Undoubtedly, photosynthesis and the subsequent burial of photosynthetically produced organic carbon are ultimately responsible for the abundant oxygen in Earth's atmosphere. But the timing and apparent suddenness of the first major increase in atmospheric oxygen is difficult to explain, in part because oxygen-producing photosynthesis seems to have evolved at least several hundred million years earlier¹³.

One possible explanation involves volcanic gases, which are chemically reducing and so

GEOCHEMISTRY

Portrait of Earth's coming of age

An analysis of geochemical data reveals a substantial change in the composition of Earth's magmas about 2.5 billion years ago, just as Earth's atmosphere and climate were also changing drastically. [SEE LETTER P.490](#)

WILLIAM M. WHITE

Geological activity is driven by two sources of energy: the radioactive decay of the elements uranium, thorium and potassium, and the initial heat generated by Earth's formation. Both decline with time. According to the standard model of terrestrial composition, Earth's energy supply is split relatively evenly between the two sources, but recent analyses^{1–3} suggest that the contribution from radioactive decay may be significantly less than previously believed. This implies that early Earth was much hotter than has been thought, and that the planet's long-term cooling is stronger. On page 490 of this issue, Keller and Schoene⁴ quantify this cooling using a large-scale analysis of the geological record.

The authors gathered about 70,000 records of igneous-rock compositions from geochemical databases to create a picture of how such compositions have changed over Earth's history. By modelling the conditions in which the magmas that formed the rocks were generated, the researchers found that the degree of mantle melting has dropped from about 35% early in

Earth's history to about 10% at present. This is consistent with other studies that have demonstrated a decline in melting over the same period. According to one such study⁵, the decline corresponds to a decrease in the mantle potential temperature — the temperature the rocks would be at if brought to the surface without loss of heat — from 1,500–1,600 °C in the late Archaean, 2.5 billion to 3.0 billion years ago, to 1,350 °C today. This rate of cooling is lower than that predicted by most compositional models, and implies that the rate of heat loss in the past must have been lower than that at present. This, in turn, suggests that plate tectonics must have operated either intermittently or more sluggishly than it does today^{2,5}.

Keller and Schoene's study also reveals a remarkable apparent discontinuity in the composition of magmas produced in both the mantle (basaltic magma) and the continental crust (granitic magmas) across the Archaean–Proterozoic boundary 2.5 billion years ago. The authors report that post-Archaean granites seem to have formed at shallower depths in the crust than Archaean granites, whereas post-Archaean basalts seem to have been generated by lesser degrees of mantle melting than

amyloid- β -induced deficits in neural function in a mouse model of Alzheimer's disease, in an APOE-dependent way¹⁰. But the effects on the BBB were not analysed in that study, leaving unanswered the question of whether increased APOE levels, changes in the lipid content of APOE particles, or other mechanisms (for instance, pericyte degeneration, as described by Bell *et al.*) contributed to the beneficial outcome of the pharmacological treatment in that mouse model.

It is tempting to speculate that pericytes fail to guard the BBB in other neurodegenerative diseases, such as Parkinson's disease and amyotrophic lateral sclerosis, in which vascular changes have been documented^{3,4}. So the provocative findings by Bell *et al.* lay down an exciting new path for future research and drug discovery regarding not only the role of APOE4 in Alzheimer's disease, but also the intriguing and fascinating interplay between the vascular and nervous systems in the development of neurodegenerative disorders. ■

Peter Carmeliet is at the *Vesalius Research Center, VIB, University of Leuven*.
Bart De Strooper is at the *VIB Center for the Biology of Disease and the Center for Human Genetics, University of Leuven, Leuven B-3000, Belgium*.
e-mails: peter.carmeliet@vib-kuleuven.be; bart.destrooper@cme.vib-kuleuven.be

1. De Strooper, B. *Physiol. Rev.* **90**, 465–494 (2010).
2. Karran, E., Mercken, M. & De Strooper, B. *Nature Rev. Drug Discov.* **10**, 698–712 (2011).
3. Quaegebeur, A., Lange, C. & Carmeliet, P. *Neuron* **71**, 406–424 (2011).
4. Storkebaum, E., Quaegebeur, A., Vikkula, M. & Carmeliet, P. *Nature Neurosci.* **14**, 1390–1397 (2011).
5. Genin, E. *et al. Mol. Psychiatry* **16**, 903–907 (2011).
6. Castellano, J. M. *et al. Sci. Transl. Med.* **3**, 89ra57 (2011).
7. Bell, R. D. *et al. Nature* **485**, 512–516 (2012).
8. Zlokovic, B. V. *Nature Rev. Neurosci.* **12**, 723–738 (2011).
9. Satoh, K., Shimokawa, H. & Berk, B. C. *Circ. J.* **74**, 2249–2256 (2010).
10. Cramer, P. E. *et al. Science* **335**, 1503–1506 (2012).

Archaean basalts. As Keller and Schoene point out⁴, the latter cannot simply reflect a drop in mantle temperature; instead it suggests a fundamental change in the way that heat moved through, and was lost by, Earth.

Geologists have long known that the Archaean–Proterozoic transition marks one of the most momentous and consequential times in Earth's history⁶ — it immediately followed the most rapid increase in Earth's crustal volume. Petrologists and geochemists have also long noted a difference in the composition of Archaean and post-Archaean continental crust that seems to be accompanied by changes in the thickness of the crust and of the lithosphere (Earth's rigid, outermost layer). A host of ideas has been suggested to explain these changes, ranging from the beginning of modern plate tectonics to an alteration in the character of subduction. One issue with Keller and Schoene's result is that there is a paucity of data from the earliest Proterozoic, but it is hard to see how that alone could explain the discontinuity. Indeed, a lull in volcanic activity in the early Proterozoic, which would explain the lack of data, has previously been noted⁷; the lull itself represents yet another puzzling aspect of the Archaean–Proterozoic transition.

Intriguingly, the Archaean–Proterozoic transition was also a time of pronounced changes at Earth's surface: between 2.4 billion and 2.2 billion years ago, the oxygen content of the atmosphere rose from a few parts per million to about 2% by volume⁸; for comparison, the modern volume fraction is around 20%. Earth also experienced its first major glaciation⁹ — one that was possibly much more severe than the Pleistocene glaciations that have occurred during the past 2 million years. Studies have shown that the oxygenation and the first glaciation occurred at almost the same time^{10,11}. Atmospheric composition exerts a strong influence on climate through the greenhouse effect, so it is not surprising that these events seem to be related. For example, the increase in oxygen might have led to a drop in levels of methane (a potent greenhouse gas), causing the greenhouse effect to crash, thereby tipping Earth into an 'ice-house' state¹². However, a consensus has yet to emerge on exactly what happened.

The fact that such marked changes seem to have occurred almost simultaneously in Earth's interior and at its surface raises the question of whether those changes are linked. Undoubtedly, photosynthesis and the subsequent burial of photosynthetically produced organic carbon are ultimately responsible for the abundant oxygen in Earth's atmosphere. But the timing and apparent suddenness of the first major increase in atmospheric oxygen is difficult to explain, in part because oxygen-producing photosynthesis seems to have evolved at least several hundred million years earlier¹³.

One possible explanation involves volcanic gases, which are chemically reducing and so

GEOCHEMISTRY

Portrait of Earth's coming of age

An analysis of geochemical data reveals a substantial change in the composition of Earth's magmas about 2.5 billion years ago, just as Earth's atmosphere and climate were also changing drastically. [SEE LETTER P.490](#)

WILLIAM M. WHITE

Geological activity is driven by two sources of energy: the radioactive decay of the elements uranium, thorium and potassium, and the initial heat generated by Earth's formation. Both decline with time. According to the standard model of terrestrial composition, Earth's energy supply is split relatively evenly between the two sources, but recent analyses^{1–3} suggest that the contribution from radioactive decay may be significantly less than previously believed. This implies that early Earth was much hotter than has been thought, and that the planet's long-term cooling is stronger. On page 490 of this issue, Keller and Schoene⁴ quantify this cooling using a large-scale analysis of the geological record.

The authors gathered about 70,000 records of igneous-rock compositions from geochemical databases to create a picture of how such compositions have changed over Earth's history. By modelling the conditions in which the magmas that formed the rocks were generated, the researchers found that the degree of mantle melting has dropped from about 35% early in

Earth's history to about 10% at present. This is consistent with other studies that have demonstrated a decline in melting over the same period. According to one such study⁵, the decline corresponds to a decrease in the mantle potential temperature — the temperature the rocks would be at if brought to the surface without loss of heat — from 1,500–1,600 °C in the late Archaean, 2.5 billion to 3.0 billion years ago, to 1,350 °C today. This rate of cooling is lower than that predicted by most compositional models, and implies that the rate of heat loss in the past must have been lower than that at present. This, in turn, suggests that plate tectonics must have operated either intermittently or more sluggishly than it does today^{2,5}.

Keller and Schoene's study also reveals a remarkable apparent discontinuity in the composition of magmas produced in both the mantle (basaltic magma) and the continental crust (granitic magmas) across the Archaean–Proterozoic boundary 2.5 billion years ago. The authors report that post-Archaean granites seem to have formed at shallower depths in the crust than Archaean granites, whereas post-Archaean basalts seem to have been generated by lesser degrees of mantle melting than

consume atmospheric oxygen. Atmospheric oxygen levels can therefore rise only when photosynthesis and carbon burial produce more oxygen than is sufficient to oxidize the volcanic gases¹³. This observation has led to a widespread suspicion that a change in the oxidation state of volcanic gases somehow controlled the timing of the rise in atmospheric oxygen. The change in volcanic-gas oxidation state could, in turn, reflect an alteration in the oxidation state of Earth's interior^{14,15}, or, alternatively, changes in factors such as the rate of volcanism or the depth at which magmas stagnate and de-gas^{7,16,17}.

Keller and Schoene add another hypothesis to this list, by proposing that the generation of granitic magmas at shallower depths was responsible for the rise in atmospheric oxygen. The chemical signature of Archaean granitic magmas suggests that they were generated at sufficient depth to leave an eclogitic residue (a mixture of the minerals garnet and pyroxene), whereas that of post-Archaean intra-crustal magmas suggests that these were generated at shallower depths and left a gabbroic residue (a mixture of olivine, pyroxene and plagioclase). Ferric iron — the oxidized form of iron — seems to be more compatible with eclogitic residues than with gabbroic ones, and so the authors speculate that an accumulation of ferric iron in eclogitic minerals may have caused the Archaean magmas, and their associated gases, to be more chemically reduced than post-Archaean ones.

The Archaean-Proterozoic boundary seems to have been when Earth came of age: a time when the planet passed from energetic, exuberant youth to more sedate middle age. But because the transition occurred an extremely long time ago, the corresponding geological record is sparse. Cause and effect have yet to be fully deduced for events that occurred around that time, but the record seems to show that processes in Earth's interior and at its surface, including the evolution of both climate and life, are intimately linked. Those events set the stage for the much later emergence of multicellular life including, ultimately, us. ■

William M. White is in the Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York 14853, USA. e-mail: wmw4@cornell.edu

1. Campbell, I. H. & O'Neill, H. St C. *Nature* **483**, 553–558 (2012).
2. White, W. M. & Phipps Morgan, J. *Mineral. Mag. (Goldschmidt Conf. Abstr.)* **75**, 1191 (2011).
3. Lyubetskaya, T. & Korenaga, J. J. *Geophys. Res.* **112**, B03211 (2007).
4. Keller, C. B. & Schoene, B. *Nature* **485**, 490–493 (2012).
5. Herzberg, C., Condie, K. & Korenaga, J. *Earth Planet. Sci. Lett.* **292**, 79–88 (2010).
6. Condie, K. C. & O'Neill, C. *Am. J. Sci.* **310**, 775–790 (2010).
7. Condie, K. C., O'Neill, C. & Aster, R. C. *Earth Planet. Sci. Lett.* **282**, 294–298 (2009).
8. Holland, H. D. *Phil. Trans. R. Soc. B* **361**, 903–915 (2006).

9. Kirschvink, J. L. *et al. Proc. Natl Acad. Sci. USA* **97**, 1400–1405 (2000).
10. Sekine, Y. *et al. Nature Commun.* **2**, 502 (2011).
11. Guo, Q. *et al. Geology* **37**, 399–402 (2009).
12. Kasting, J. F. & Ono, S. *Phil. Trans. R. Soc. B* **361**, 917–929 (2006).
13. Lyons, T. W. & Reinhard, C. T. *Nature* **478**, 194–195 (2011).

14. Kasting, J. F., Eggler, D. H. & Raeburn, S. P. *J. Geol.* **101**, 245–257 (1993).
15. Kump, L. R., Kasting, J. F. & Barley, M. E. *Geochim. Geophys. Geosyst.* <http://dx.doi.org/10.1029/2000GC000114> (2001).
16. Kump, L. R. & Barley, M. E. *Nature* **448**, 1033–1036 (2007).
17. Gaillard, F., Scaillet, B. & Arndt, N. T. *Nature* **478**, 229–232 (2011).

NEUROSCIENCE

Crystal-clear brains

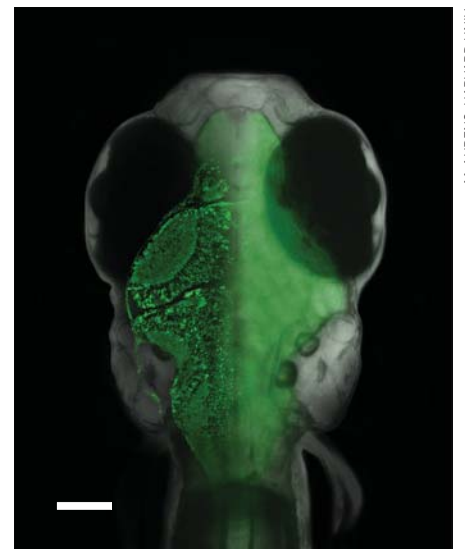
An ingenious technique allows the monitoring of brain-wide patterns of neuronal activity in a vertebrate at the cellular level, while the animal interacts with a virtual environment. SEE ARTICLE P.471

JOSEPH R. FETCHO

Neurobiologists studying brains face an enormous problem. They must deal with massive numbers of neurons — perhaps as many as 100 billion in a human brain — that interact with one another to produce the electrical activity that drives behaviour. Moreover, both the interactions and the activity change constantly as animals learn to modify their behaviour as a result of experience. How can the problem of interpreting such dynamics be addressed? One initial step is to monitor the activity of every individual neuron in the brain to reveal activity patterns during a particular behaviour, and to observe how those patterns change with experience. This is exactly what Ahrens *et al.*¹ have done for the first time using a vertebrate animal model. On page 471 of this issue, the authors describe how they imaged the activity of individual neurons throughout the brain of zebrafish larvae while the animals interacted with a changing virtual environment and adjusted their behaviour accordingly.

Most animals are opaque, so obtaining images of their brain with the resolution needed to see individual neurons (tens of micrometres or less) might seem impossible. But Ahrens and colleagues used a small, transparent vertebrate that is amenable to genetic manipulation — the larval zebrafish (Fig. 1). To detect the nerve cells that were active at any given moment, the authors introduced into the animal a gene coding for a fluorescent calcium indicator, which labels brain neurons and becomes brighter as calcium enters electrically active cells^{2–4}. This, combined with advanced optical methods (such as two-photon microscopy), allowed the researchers to see the active neurons light up anywhere in the brain.

The authors set out to study a key problem in neuroscience: how do patterns of neuronal activity change throughout the brain as an animal modifies its behaviour when it learns something new? Fish adjust how strongly they



M. AHRENS, HARVARD UNIV.

Figure 1 | Illuminating behaviour. Ahrens *et al.*¹ monitored neuronal function throughout the brain of zebrafish larvae while the animals performed a learning task. Overlaid on an image of the fish is an image of a slice of its brain, as obtained with a laser scanning microscope, in which neurons are labelled with a green fluorescent calcium-indicator protein. Scale bar, 100 micrometres.

activate their muscles by using visual feedback about how quickly the world is passing by as an indication of the speed at which they are moving. To facilitate imaging while fish learned to make these adjustments, Ahrens *et al.* immobilized zebrafish larvae by injecting them with a paralyzing agent, and put them into a simulated virtual environment. Using computer monitors, the researchers controlled the rate at which the virtual world seemed to move past when the fish tried to move. The virtual scenes changed according to the animals' intended movements, which were inferred by the detection of electrical signals from nerves connected to the muscles that would normally produce the movements.

By changing how fast the world seemed

consume atmospheric oxygen. Atmospheric oxygen levels can therefore rise only when photosynthesis and carbon burial produce more oxygen than is sufficient to oxidize the volcanic gases¹³. This observation has led to a widespread suspicion that a change in the oxidation state of volcanic gases somehow controlled the timing of the rise in atmospheric oxygen. The change in volcanic-gas oxidation state could, in turn, reflect an alteration in the oxidation state of Earth's interior^{14,15}, or, alternatively, changes in factors such as the rate of volcanism or the depth at which magmas stagnate and de-gas^{7,16,17}.

Keller and Schoene add another hypothesis to this list, by proposing that the generation of granitic magmas at shallower depths was responsible for the rise in atmospheric oxygen. The chemical signature of Archaean granitic magmas suggests that they were generated at sufficient depth to leave an eclogitic residue (a mixture of the minerals garnet and pyroxene), whereas that of post-Archaean intra-crustal magmas suggests that these were generated at shallower depths and left a gabbroic residue (a mixture of olivine, pyroxene and plagioclase). Ferric iron — the oxidized form of iron — seems to be more compatible with eclogitic residues than with gabbroic ones, and so the authors speculate that an accumulation of ferric iron in eclogitic minerals may have caused the Archaean magmas, and their associated gases, to be more chemically reduced than post-Archaean ones.

The Archaean-Proterozoic boundary seems to have been when Earth came of age: a time when the planet passed from energetic, exuberant youth to more sedate middle age. But because the transition occurred an extremely long time ago, the corresponding geological record is sparse. Cause and effect have yet to be fully deduced for events that occurred around that time, but the record seems to show that processes in Earth's interior and at its surface, including the evolution of both climate and life, are intimately linked. Those events set the stage for the much later emergence of multicellular life including, ultimately, us. ■

William M. White is in the Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York 14853, USA. e-mail: wmw4@cornell.edu

1. Campbell, I. H. & O'Neill, H. St C. *Nature* **483**, 553–558 (2012).
2. White, W. M. & Phipps Morgan, J. *Mineral. Mag. (Goldschmidt Conf. Abstr.)* **75**, 1191 (2011).
3. Lyubetskaya, T. & Korenaga, J. J. *Geophys. Res.* **112**, B03211 (2007).
4. Keller, C. B. & Schoene, B. *Nature* **485**, 490–493 (2012).
5. Herzberg, C., Condie, K. & Korenaga, J. *Earth Planet. Sci. Lett.* **292**, 79–88 (2010).
6. Condie, K. C. & O'Neill, C. *Am. J. Sci.* **310**, 775–790 (2010).
7. Condie, K. C., O'Neill, C. & Aster, R. C. *Earth Planet. Sci. Lett.* **282**, 294–298 (2009).
8. Holland, H. D. *Phil. Trans. R. Soc. B* **361**, 903–915 (2006).

9. Kirschvink, J. L. *et al. Proc. Natl Acad. Sci. USA* **97**, 1400–1405 (2000).
10. Sekine, Y. *et al. Nature Commun.* **2**, 502 (2011).
11. Guo, Q. *et al. Geology* **37**, 399–402 (2009).
12. Kasting, J. F. & Ono, S. *Phil. Trans. R. Soc. B* **361**, 917–929 (2006).
13. Lyons, T. W. & Reinhard, C. T. *Nature* **478**, 194–195 (2011).

14. Kasting, J. F., Egger, D. H. & Raeburn, S. P. *J. Geol.* **101**, 245–257 (1993).
15. Kump, L. R., Kasting, J. F. & Barley, M. E. *Geochim. Geophys. Geosyst.* <http://dx.doi.org/10.1029/2000GC000114> (2001).
16. Kump, L. R. & Barley, M. E. *Nature* **448**, 1033–1036 (2007).
17. Gaillard, F., Scaillet, B. & Arndt, N. T. *Nature* **478**, 229–232 (2011).

NEUROSCIENCE

Crystal-clear brains

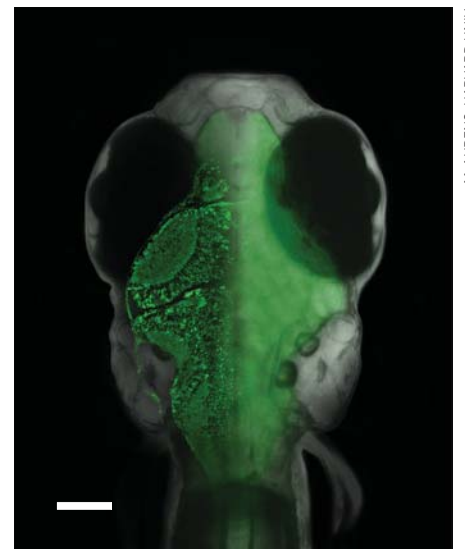
An ingenious technique allows the monitoring of brain-wide patterns of neuronal activity in a vertebrate at the cellular level, while the animal interacts with a virtual environment. SEE ARTICLE P.471

JOSEPH R. FETCHO

Neurobiologists studying brains face an enormous problem. They must deal with massive numbers of neurons — perhaps as many as 100 billion in a human brain — that interact with one another to produce the electrical activity that drives behaviour. Moreover, both the interactions and the activity change constantly as animals learn to modify their behaviour as a result of experience. How can the problem of interpreting such dynamics be addressed? One initial step is to monitor the activity of every individual neuron in the brain to reveal activity patterns during a particular behaviour, and to observe how those patterns change with experience. This is exactly what Ahrens *et al.*¹ have done for the first time using a vertebrate animal model. On page 471 of this issue, the authors describe how they imaged the activity of individual neurons throughout the brain of zebrafish larvae while the animals interacted with a changing virtual environment and adjusted their behaviour accordingly.

Most animals are opaque, so obtaining images of their brain with the resolution needed to see individual neurons (tens of micrometres or less) might seem impossible. But Ahrens and colleagues used a small, transparent vertebrate that is amenable to genetic manipulation — the larval zebrafish (Fig. 1). To detect the nerve cells that were active at any given moment, the authors introduced into the animal a gene coding for a fluorescent calcium indicator, which labels brain neurons and becomes brighter as calcium enters electrically active cells^{2–4}. This, combined with advanced optical methods (such as two-photon microscopy), allowed the researchers to see the active neurons light up anywhere in the brain.

The authors set out to study a key problem in neuroscience: how do patterns of neuronal activity change throughout the brain as an animal modifies its behaviour when it learns something new? Fish adjust how strongly they



M. AHRENS, HARVARD UNIV.

Figure 1 | Illuminating behaviour. Ahrens *et al.*¹ monitored neuronal function throughout the brain of zebrafish larvae while the animals performed a learning task. Overlaid on an image of the fish is an image of a slice of its brain, as obtained with a laser scanning microscope, in which neurons are labelled with a green fluorescent calcium-indicator protein. Scale bar, 100 micrometres.

activate their muscles by using visual feedback about how quickly the world is passing by as an indication of the speed at which they are moving. To facilitate imaging while fish learned to make these adjustments, Ahrens *et al.* immobilized zebrafish larvae by injecting them with a paralyzing agent, and put them into a simulated virtual environment. Using computer monitors, the researchers controlled the rate at which the virtual world seemed to move past when the fish tried to move. The virtual scenes changed according to the animals' intended movements, which were inferred by the detection of electrical signals from nerves connected to the muscles that would normally produce the movements.

By changing how fast the world seemed

to move past when the paralysed animals attempted to activate their muscles, the authors tricked the fish into thinking that their muscles were suddenly weaker or stronger than normal. The fish learned to compensate for the changes by altering the strength of muscle activation to allow them to control their movements in the changing virtual environment.

The authors found that, even in the relatively simple brain of the larval zebrafish, the neurons that showed increased or decreased activity during learning were widespread. As expected, some of these cells were located in regions that receive information from the eyes as well as in others involved in producing the movements. Other brain areas that changed activity included the cerebellum and the inferior olive, which are crucial for complex control of movements, and the pallium in the front of the brain, a higher-level processing region. Of note, these changes in activity patterns, although extensive, are probably an underestimate of the extent of the changes that occur during learning — although the researchers' approach is the most powerful available, it is prone to missing small changes in the activity of individual neurons.

Nonetheless, Ahrens and colleagues' work faces head-on the problem of the complex ties between the brain and behaviour. Although behaviour is known to emerge from the activity of, and the interactions among, neurons across many brain regions, neurobiologists typically focus their research on just one of these regions at a time because of the technical difficulties associated with brain-wide studies at the cellular level. Mapping activity patterns at the single-cell level throughout a brain while the animal is learning is a major step towards a more integrated view of how neuronal circuits, in their entirety, drive behaviour.

The authors' study is a remarkable achievement, but documenting the patterns of activity everywhere in the brain is just the beginning of the task ahead. Many important details about the neurons whose activity changed during Ahrens and colleagues' experiments are unknown. We need not only a broad picture of the activity at the cellular level as provided by the authors' work, but also information about whether the neurons excite or inhibit other cells and how they connect to other neurons in the brain to form the circuits that drive behaviour. This is the focus of 'connectomics', an effort to map all of the connections in the brain⁵. Once neuronal activity, and both structural and functional connectivity, are mapped, models of the way in which circuits throughout the brain produce and modify behaviour must be formulated and tested. A powerful technique for this purpose is optogenetics⁶, which allows neurons to be turned on and off with light during behaviour to test the contributions of different sets of cells — an approach that is especially applicable to a transparent animal. Therefore, all of

the necessary tools for mapping activity and connectivity, and for testing neuronal function on a large scale, are now at hand and promise to provide increasingly complete pictures of behaviour-generating circuits throughout the brain.

The larval zebrafish will probably remain at the forefront of these efforts because of the powerful combination of optical, genetic and behavioural approaches that can be used to study it⁷. These larval fish show most of the behaviours typical of adult vertebrates, although some aspects of their behaviour are not as highly refined as that of adult fish or mammals. Most importantly, in no other vertebrate animal model is it possible to accomplish cellular-level-resolution imaging of the entire brain. The tools to map connectivity and perturb neuronal activity are also easily applied to this small, transparent fish, making it the best hope among vertebrate models for revealing how neurons throughout the brain

interact to produce behaviour^{8–10}. With the help of a tiny fish, a long-standing, but distant, goal of neuroscience is finally moving into view. ■

Joseph R. Fetcho is in the Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA. e-mail: jrfetcho@cornell.edu

1. Ahrens, M. B. *et al.* *Nature* **485**, 471–477 (2012).
2. Miyawaki, A. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 2135–2140 (1999).
3. Higashijima, S. *et al.* *J. Neurophysiol.* **90**, 3986–3997 (2003).
4. Looger, L. & Griesbeck, O. *Curr. Opin. Neurobiol.* **22**, 18–23 (2012).
5. Lichtman, J. W. *et al.* *Nature Rev. Neurosci.* **9**, 417–422 (2008).
6. Boyden, E. S. *et al.* *Nature Neurosci.* **8**, 1263–1268 (2005).
7. Fetcho, J. R. & Liu, K. S. *Ann. N. Y. Acad. Sci.* **860**, 333–345 (1998).
8. Douglass, A. D. *Curr. Biol.* **18**, 1133–1137 (2008).
9. Kinkhabwala, A. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 1164–1169 (2011).
10. Koyama, M. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 1170–1175 (2011).

PARTICLE PHYSICS

A reminder of the beauty we know

The latest measurements of the mass of the W boson, one of two elementary particles that mediate the weak nuclear force, are a powerful reminder of the profound beauty in the standard model of particle physics.

JONATHAN BUTTERWORTH

The most precise measurement so far of the mass of the W boson is reported by researchers in the CDF Collaboration (Aaltonen *et al.*¹) at the Tevatron collider, Fermilab, near Chicago, in *Physical Review Letters*. The W boson is the particle that carries the weak nuclear force, which is responsible for radioactive β -decay. The particle decays in less than one billion-trillionth of a second, and many of the things that it decays to are hard to detect. The measurement of its mass is an experimental tour de force, a consistency test for the standard model of particle physics, and a pointer that will aid the discovery of new phenomena at the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

The 'standard model' is the rather prosaic name given to the collection of quantum field theories that describes the Universe at the shortest distances and highest energies. You will often hear particle physicists hoping for physics 'beyond the standard model' because an anomalous experimental result might help to resolve some of the issues left unexplained by the model. These problems — including

why the Universe contains more matter than antimatter, and where gravity fits in — are discussed so often that it is easy to forget how accurate and economical the standard model is.

In the model, the fundamental forces are carried by particles known as gauge bosons. The photon, for example, is the gauge boson that carries the electromagnetic force, whereas the W and Z bosons mediate the weak force. Bosons are particles that have an integer quantum of intrinsic angular momentum, otherwise known as spin. 'Gauge' means that the particles are generated by mathematical symmetries.

Symmetry is a central feature in the standard model. For the model to work in its most symmetric form, the W and Z bosons should, like the photon, have zero mass. This is not the case. Yet without symmetry, the standard model loses all predictive power — a finding² that earned Gerardus 't Hooft and Martinus Veltman the Nobel Prize in Physics in 1999.

The way out of this impasse could be supplied by the discovery of the Higgs boson. Its existence would indicate that symmetry remains in the standard model at high

to move past when the paralysed animals attempted to activate their muscles, the authors tricked the fish into thinking that their muscles were suddenly weaker or stronger than normal. The fish learned to compensate for the changes by altering the strength of muscle activation to allow them to control their movements in the changing virtual environment.

The authors found that, even in the relatively simple brain of the larval zebrafish, the neurons that showed increased or decreased activity during learning were widespread. As expected, some of these cells were located in regions that receive information from the eyes as well as in others involved in producing the movements. Other brain areas that changed activity included the cerebellum and the inferior olive, which are crucial for complex control of movements, and the pallium in the front of the brain, a higher-level processing region. Of note, these changes in activity patterns, although extensive, are probably an underestimate of the extent of the changes that occur during learning — although the researchers' approach is the most powerful available, it is prone to missing small changes in the activity of individual neurons.

Nonetheless, Ahrens and colleagues' work faces head-on the problem of the complex ties between the brain and behaviour. Although behaviour is known to emerge from the activity of, and the interactions among, neurons across many brain regions, neurobiologists typically focus their research on just one of these regions at a time because of the technical difficulties associated with brain-wide studies at the cellular level. Mapping activity patterns at the single-cell level throughout a brain while the animal is learning is a major step towards a more integrated view of how neuronal circuits, in their entirety, drive behaviour.

The authors' study is a remarkable achievement, but documenting the patterns of activity everywhere in the brain is just the beginning of the task ahead. Many important details about the neurons whose activity changed during Ahrens and colleagues' experiments are unknown. We need not only a broad picture of the activity at the cellular level as provided by the authors' work, but also information about whether the neurons excite or inhibit other cells and how they connect to other neurons in the brain to form the circuits that drive behaviour. This is the focus of 'connectomics', an effort to map all of the connections in the brain⁵. Once neuronal activity, and both structural and functional connectivity, are mapped, models of the way in which circuits throughout the brain produce and modify behaviour must be formulated and tested. A powerful technique for this purpose is optogenetics⁶, which allows neurons to be turned on and off with light during behaviour to test the contributions of different sets of cells — an approach that is especially applicable to a transparent animal. Therefore, all of

the necessary tools for mapping activity and connectivity, and for testing neuronal function on a large scale, are now at hand and promise to provide increasingly complete pictures of behaviour-generating circuits throughout the brain.

The larval zebrafish will probably remain at the forefront of these efforts because of the powerful combination of optical, genetic and behavioural approaches that can be used to study it⁷. These larval fish show most of the behaviours typical of adult vertebrates, although some aspects of their behaviour are not as highly refined as that of adult fish or mammals. Most importantly, in no other vertebrate animal model is it possible to accomplish cellular-level-resolution imaging of the entire brain. The tools to map connectivity and perturb neuronal activity are also easily applied to this small, transparent fish, making it the best hope among vertebrate models for revealing how neurons throughout the brain

interact to produce behaviour^{8–10}. With the help of a tiny fish, a long-standing, but distant, goal of neuroscience is finally moving into view. ■

Joseph R. Fetcho is in the Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA. e-mail: jfetcho@cornell.edu

1. Ahrens, M. B. *et al.* *Nature* **485**, 471–477 (2012).
2. Miyawaki, A. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 2135–2140 (1999).
3. Higashijima, S. *et al.* *J. Neurophysiol.* **90**, 3986–3997 (2003).
4. Looger, L. & Griesbeck, O. *Curr. Opin. Neurobiol.* **22**, 18–23 (2012).
5. Lichtman, J. W. *et al.* *Nature Rev. Neurosci.* **9**, 417–422 (2008).
6. Boyden, E. S. *et al.* *Nature Neurosci.* **8**, 1263–1268 (2005).
7. Fetcho, J. R. & Liu, K. S. *Ann. N. Y. Acad. Sci.* **860**, 333–345 (1998).
8. Douglass, A. D. *Curr. Biol.* **18**, 1133–1137 (2008).
9. Kinkhabwala, A. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 1164–1169 (2011).
10. Koyama, M. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 1170–1175 (2011).

PARTICLE PHYSICS

A reminder of the beauty we know

The latest measurements of the mass of the W boson, one of two elementary particles that mediate the weak nuclear force, are a powerful reminder of the profound beauty in the standard model of particle physics.

JONATHAN BUTTERWORTH

The most precise measurement so far of the mass of the W boson is reported by researchers in the CDF Collaboration (Aaltonen *et al.*¹) at the Tevatron collider, Fermilab, near Chicago, in *Physical Review Letters*. The W boson is the particle that carries the weak nuclear force, which is responsible for radioactive β -decay. The particle decays in less than one billion-trillionth of a second, and many of the things that it decays to are hard to detect. The measurement of its mass is an experimental tour de force, a consistency test for the standard model of particle physics, and a pointer that will aid the discovery of new phenomena at the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

The 'standard model' is the rather prosaic name given to the collection of quantum field theories that describes the Universe at the shortest distances and highest energies. You will often hear particle physicists hoping for physics 'beyond the standard model' because an anomalous experimental result might help to resolve some of the issues left unexplained by the model. These problems — including

why the Universe contains more matter than antimatter, and where gravity fits in — are discussed so often that it is easy to forget how accurate and economical the standard model is.

In the model, the fundamental forces are carried by particles known as gauge bosons. The photon, for example, is the gauge boson that carries the electromagnetic force, whereas the W and Z bosons mediate the weak force. Bosons are particles that have an integer quantum of intrinsic angular momentum, otherwise known as spin. 'Gauge' means that the particles are generated by mathematical symmetries.

Symmetry is a central feature in the standard model. For the model to work in its most symmetric form, the W and Z bosons should, like the photon, have zero mass. This is not the case. Yet without symmetry, the standard model loses all predictive power — a finding² that earned Gerardus 't Hooft and Martinus Veltman the Nobel Prize in Physics in 1999.

The way out of this impasse could be supplied by the discovery of the Higgs boson. Its existence would indicate that symmetry remains in the standard model at high



R. HAHN/FERMI LAB

Figure 1 | The Tevatron collider. Aaltonen *et al.*¹ have used the Tevatron — a 6.3-kilometre-long circular particle accelerator and collider — at Fermilab in Batavia, Illinois, to measure the mass of the W boson to high precision.

energies, even though it is manifestly broken in the low-energy Universe in which we live.

The non-zero mass of the W boson is intimately connected with the Higgs boson, with the origins of mass in general and with our understanding of physics in terms of quantum field theories. It is a quantity well worth measuring precisely — just as the Tevatron experimenters¹ have done.

This was a hugely challenging analysis. The Tevatron (Fig. 1), which recently ceased operation, was a high-energy collider that stored protons and antiprotons, accelerated them to high energies — almost 1 teraelectronvolt (10^{12} eV) — and forced them into head-on collisions. The energy and frequency of the collisions were sufficient to produce large numbers of W bosons. W bosons decay rapidly, and their decay products could generally be detected in the Collider Detector at Fermilab (CDF), or in the rival detector D0, which is also located at the Tevatron. In the CDF analysis, the researchers used 1,094,834 W-boson decays to measure the W-boson mass.

The W boson can decay in many different ways, but those decays that produce an electron or a muon — a short-lived particle similar to the electron — are the most useful for measuring the W boson's mass because electrons and muons can be reliably detected. However, an electrically neutral particle called a neutrino that is hard to detect is also produced in these decay events. This is problematic because the neutrino's momentum is needed to determine the W boson's mass; however, this momentum can be deduced only indirectly from an analysis of all the other particles produced in the decay event.

The CDF is a cylinder constructed such that proton and antiproton beams enter at either end and collide in the centre. Although the neutrino cannot be detected, its presence — and the component of its momentum transverse to the beam — can be deduced by applying the law of conservation of momentum to all the other particles produced in the collision. In addition to the electron and the muon, this includes composite particles known as hadrons, which are generated when elementary particles called quarks and gluons are scattered from the colliding particles and then combine.

Detailed analyses of all of these components led the CDF Collaboration to obtain a value for the W boson's mass of 80,387 MeV with an error of 19 MeV, a precision of about two parts in 10,000. This value is consistent with that obtained from an experiment performed with the D0 detector, which found³ a mass of 80,367 MeV with an error of 26 MeV.

The W boson and the top quark, the heaviest of all known elementary particles, contribute to many particle-production and scattering processes that have been accurately measured in particle-physics experiments. In these processes, the particles enter quantum loops as virtual particles with fleeting existence but measurable effect — at least, if the measurement is precise enough. If it exists, the Higgs boson must also appear in these loops. By combining these measurements with their value of the W boson's mass, the authors were able to conduct a precise test of the symmetry structure of the standard model.

Knowledge of the W boson's mass has imposed limits on the range of possible mass values for the Higgs boson. This range has

been further curtailed^{4,5} by data from the direct searches for the Higgs at the LHC. And yet there is still a region of overlap. If the hints seen at the LHC do turn out to be the Higgs, then the particle's mass is consistent with that inferred from standard-model calculations, using the W boson's mass, of an array of particle-physics processes. This consistency is built into the quantum loops and symmetries of the standard model. A theorist, or even a mathematician, might call this a highly non-trivial consistency test of the theory. I call it beautiful. ■

Jonathan Butterworth is in the Department of Physics and Astronomy, University College London, London WC1E 6BT, UK.
e-mail: j.butterworth@ucl.ac.uk

1. Aaltonen, T. *et al.* *Phys. Rev. Lett.* **108**, 151803 (2012).
2. 't Hooft, G. & Veltman, M. *Nucl. Phys. B* **44**, 189–213 (1972).
3. Abazov, V. M. *et al.* *Phys. Rev. Lett.* **108**, 151804 (2012).
4. Aad, G. *et al.* *Phys. Lett. B* **710**, 49–66 (2012).
5. Chatrchyan, S. *et al.* *Phys. Lett. B* **710**, 26–48 (2012).

ASTROPHYSICS

Startling superflares

Stars that are just like our Sun have flares more than a million times more energetic than the biggest flare ever seen on the Sun. The Kepler satellite has allowed these superflares to be studied in detail for the first time. [SEE LETTER P.478](#)

BRADLEY E. SCHAEFER

A superflare on a Sun-like star is a brightening that has an energy of from 10^{33} to more than 10^{39} erg and lasts from minutes to days. The Sun has frequent flares that are caused by magnetic effects above sunspots, regions that are cooler than the Sun's typical surface temperature. However, the largest flare ever observed¹ on the Sun — the 1859 Carrington event — had a total energy of about 10^{32} erg. With Sun-like stars being the epitome of constancy, it is startling, evocative

and exciting that they can have superflares as energetic as 10^{39} erg. On page 478 of this issue, Maehara *et al.*² report the emissions from 365 superflares, measured by the awesome Kepler satellite, which was launched in 2009*.

Over the past 120 years, four dozen superflares have been reported in the literature^{3–5}. But these events were always ignored as isolated anomalies. Only in 1989 were all these reports put together in recognition that the events represent a coherent phenomenon,

*This article and the paper² under discussion were published online on 16 May 2012.



R. HAHN/FERMI-LAB

Figure 1 | The Tevatron collider. Aaltonen *et al.*¹ have used the Tevatron — a 6.3-kilometre-long circular particle accelerator and collider — at Fermilab in Batavia, Illinois, to measure the mass of the W boson to high precision.

energies, even though it is manifestly broken in the low-energy Universe in which we live.

The non-zero mass of the W boson is intimately connected with the Higgs boson, with the origins of mass in general and with our understanding of physics in terms of quantum field theories. It is a quantity well worth measuring precisely — just as the Tevatron experimenters¹ have done.

This was a hugely challenging analysis. The Tevatron (Fig. 1), which recently ceased operation, was a high-energy collider that stored protons and antiprotons, accelerated them to high energies — almost 1 teraelectronvolt (10^{12} eV) — and forced them into head-on collisions. The energy and frequency of the collisions were sufficient to produce large numbers of W bosons. W bosons decay rapidly, and their decay products could generally be detected in the Collider Detector at Fermilab (CDF), or in the rival detector D0, which is also located at the Tevatron. In the CDF analysis, the researchers used 1,094,834 W-boson decays to measure the W-boson mass.

The W boson can decay in many different ways, but those decays that produce an electron or a muon — a short-lived particle similar to the electron — are the most useful for measuring the W boson's mass because electrons and muons can be reliably detected. However, an electrically neutral particle called a neutrino that is hard to detect is also produced in these decay events. This is problematic because the neutrino's momentum is needed to determine the W boson's mass; however, this momentum can be deduced only indirectly from an analysis of all the other particles produced in the decay event.

The CDF is a cylinder constructed such that proton and antiproton beams enter at either end and collide in the centre. Although the neutrino cannot be detected, its presence — and the component of its momentum transverse to the beam — can be deduced by applying the law of conservation of momentum to all the other particles produced in the collision. In addition to the electron and the muon, this includes composite particles known as hadrons, which are generated when elementary particles called quarks and gluons are scattered from the colliding particles and then combine.

Detailed analyses of all of these components led the CDF Collaboration to obtain a value for the W boson's mass of 80,387 MeV with an error of 19 MeV, a precision of about two parts in 10,000. This value is consistent with that obtained from an experiment performed with the D0 detector, which found³ a mass of 80,367 MeV with an error of 26 MeV.

The W boson and the top quark, the heaviest of all known elementary particles, contribute to many particle-production and scattering processes that have been accurately measured in particle-physics experiments. In these processes, the particles enter quantum loops as virtual particles with fleeting existence but measurable effect — at least, if the measurement is precise enough. If it exists, the Higgs boson must also appear in these loops. By combining these measurements with their value of the W boson's mass, the authors were able to conduct a precise test of the symmetry structure of the standard model.

Knowledge of the W boson's mass has imposed limits on the range of possible mass values for the Higgs boson. This range has

been further curtailed^{4,5} by data from the direct searches for the Higgs at the LHC. And yet there is still a region of overlap. If the hints seen at the LHC do turn out to be the Higgs, then the particle's mass is consistent with that inferred from standard-model calculations, using the W boson's mass, of an array of particle-physics processes. This consistency is built into the quantum loops and symmetries of the standard model. A theorist, or even a mathematician, might call this a highly non-trivial consistency test of the theory. I call it beautiful. ■

Jonathan Butterworth is in the Department of Physics and Astronomy, University College London, London WC1E 6BT, UK.
e-mail: j.butterworth@ucl.ac.uk

1. Aaltonen, T. *et al.* *Phys. Rev. Lett.* **108**, 151803 (2012).
2. 't Hooft, G. & Veltman, M. *Nucl. Phys. B* **44**, 189–213 (1972).
3. Abazov, V. M. *et al.* *Phys. Rev. Lett.* **108**, 151804 (2012).
4. Aad, G. *et al.* *Phys. Lett. B* **710**, 49–66 (2012).
5. Chatrchyan, S. *et al.* *Phys. Lett. B* **710**, 26–48 (2012).

ASTROPHYSICS

Startling superflares

Stars that are just like our Sun have flares more than a million times more energetic than the biggest flare ever seen on the Sun. The Kepler satellite has allowed these superflares to be studied in detail for the first time. [SEE LETTER P.478](#)

BRADLEY E. SCHAEFER

A superflare on a Sun-like star is a brightening that has an energy of from 10^{33} to more than 10^{39} erg and lasts from minutes to days. The Sun has frequent flares that are caused by magnetic effects above sunspots, regions that are cooler than the Sun's typical surface temperature. However, the largest flare ever observed¹ on the Sun — the 1859 Carrington event — had a total energy of about 10^{32} erg. With Sun-like stars being the epitome of constancy, it is startling, evocative

and exciting that they can have superflares as energetic as 10^{39} erg. On page 478 of this issue, Maehara *et al.*² report the emissions from 365 superflares, measured by the awesome Kepler satellite, which was launched in 2009*.

Over the past 120 years, four dozen superflares have been reported in the literature^{3–5}. But these events were always ignored as isolated anomalies. Only in 1989 were all these reports put together in recognition that the events represent a coherent phenomenon,

*This article and the paper² under discussion were published online on 16 May 2012.

extending to all types of normal stars^{3–5}, and the name superflares was coined to distinguish them from ordinary flares such as those that happen on the Sun⁵.

Superflares occur on single, middle-aged stars that are rotating slowly and are powered by the fusion of hydrogen in their core. Such stars, which are technically known as main-sequence stars of spectral type F8 to G8, include the closest known ‘twins’ of our Sun. The similarity of superflares to solar flares suggests that superflares arise from magnetic effects. However, the now-default model^{6–9} for these events involves a magnetic field that connects the star to an orbiting ‘hot Jupiter’ — a planet that has a mass comparable to, or larger than, that of Jupiter but that is much closer to its host star than Jupiter is to the Sun (Fig. 1).

The limitation of pre-Kepler observations is that they are an inhomogeneous collection of data. Some data are from X-ray satellites, whereas others are from spectroscopic observations, multicolour photometry and even visual observations. Such an assorted data set makes it impossible to calculate rates of occurrence, or to seek correlations between observed features or do other statistical analyses. With the average recurrence timescale for the brightest events being much longer than a decade, it was unfeasible to set up any realistic programme to observe superflares, and so the field stagnated.

Maehara and colleagues² now break this deadlock by analysing data from the Kepler satellite, which provides continuous monitoring of the brightness of more than 100,000 stars for years on end with an accuracy of 10 parts per million. The authors have looked at some 83,000 G-type main-sequence stars and provide light curves — plots showing the evolution of an object’s brightness over time — for 365 superflares on 148 stars. Suddenly, we have a wealth of data, we can do statistical analyses, and we know exactly which stars to watch.

The Kepler superflares have durations of 1–12 hours, brightness increases of 0.1–30% and total energies of 10^{33} – 10^{36} erg. A typical star in the Kepler data set has 10^{35} -erg flares every 100 days. Interestingly, Maehara *et al.* find that the observed superflares all occur on stars that have large starspots, as evidenced by quasi-periodic modulation of the star’s quiescent brightness. This finding crucially ties the superflares to starspots and hence to magnetic fields. Another insight is that the superflare stars apparently have no planetary transits — that is, no planets have been found to pass in front of them. However, roughly 10% of the superflare stars should have such transits if they are all associated with hot Jupiters. So the mechanism underlying superflares remains unclear.

With Maehara and colleagues’ results, theorists have a rich field to investigate. And

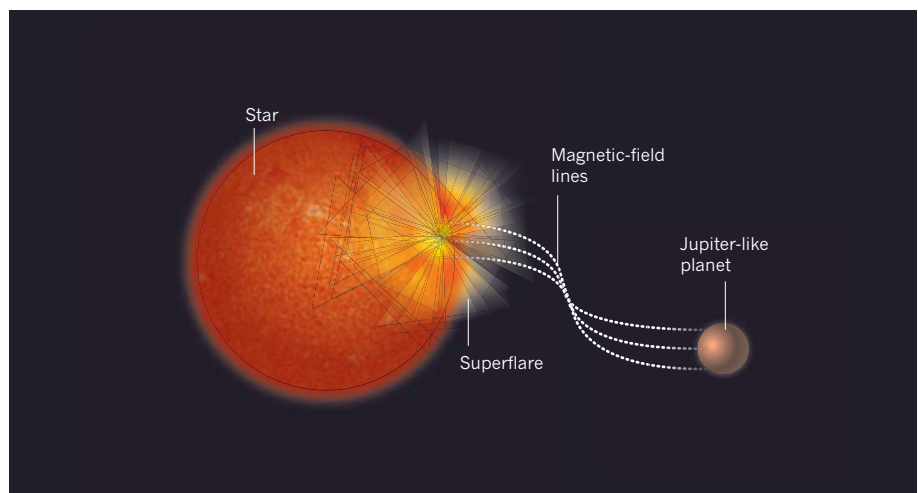


Figure 1 | Magnetic connection. One idea to explain the superflares observed by Maehara *et al.*² invokes the presence of intense magnetic fields that connect the star with a Jupiter-like planet in very close orbit around the star. The magnetic-field lines will become twisted and amplified by the orbital motion of the planet, and at some time the lines will be strained and twisted to the point of breaking. The broken lines will accelerate particles to very high energy and release this energy in an explosive event, similar to what happens in ordinary solar flares seen on the Sun.

I can think of many new paths that have been opened up for observers. Measurement of the radial velocity of the superflare stars might reveal Jupiter-like planets. The brightest of the superflare stars are bright enough to allow the stars’ magnetic fields to be determined. High-resolution spectra of the most active stars (which have superflares every 9 days) could be obtained with a view to detecting changes in the shape of the calcium spectral lines (the H and K lines) simultaneous with a Kepler superflare. Such changes would provide information about the velocity, temperature and energy of the superflares. The Kepler data could be further scrutinized to investigate whether flares occur in preferred phases of the star’s brightness modulation and whether the energies and time intervals of successive flares are correlated. Maehara and colleagues’ work could also be extended to stars of all types. Indeed, this task of searching for flares is perfect for pursuits such as the Planet Hunters programme of citizen science, which is part of the Zooniverse project¹⁰.

The possibility that the Sun has superflares is not realized. Historical and geophysical records show that the Sun has not had any superflares in the past two millennia, and no superflares with more than roughly 10^{36} erg for perhaps a billion years⁵. Maehara *et al.* show that only 0.2% of Sun-like stars have superflares, so it is unlikely that the Sun has such events. With their average rate of occurrence (once every 100 days for 10^{35} -erg flares) and their observed size distribution (with a power-law index of roughly -2.0), the expected frequency of 10^{32} -erg flares on all superflare stars should be very high. In stark contrast to this, the Sun has one 10^{32} -erg event roughly every 450 years^{1,11} and so is completely different from superflare stars. This is all readily understood within the

default model, because the Sun does not have a hot Jupiter.

Superflares have implications far beyond being just a challenge for stellar physics. If a superflare’s energy is linked to the orbital energy of a hot Jupiter, then three events a year on the star would make its planetary companion spiral in towards it on a timescale of a billion years. The huge energy output of superflares could make any planets around the star uninhabitable for far-future human colonization, and astrobiologists will have to consider the effect of the superflares on possible alien life. Superflares might provide the high-energy radiation required to create organic molecules, so perhaps superflare systems are a good place to look for alien life that has evolved to avoid the effects of the huge flares. ■

Bradley E. Schaefer is in the Department of Physics and Astronomy, Louisiana State University, Baton Rouge, Louisiana 70803, USA.

e-mail: schaefer@lsu.edu

1. Tsurutani, B. T., Gonzalez, W. D., Lakhina, G. S. & Alex, S. J. *Geophys. Res.* **108**, 1268 (2003).
2. Maehara, H. *et al.* *Nature* **485**, 478–481 (2012).
3. Schaefer, B. E. *Astrophys. J.* **337**, 927–933 (1989).
4. Schaefer, B. E. *Astrophys. J.* **366**, L39–L42 (1991).
5. Schaefer, B. E., King, J. R. & Deliyannis, C. P. *Astrophys. J.* **529**, 1026–1030 (2000).
6. Rubenstein, E. P. & Schaefer, B. E. *Astrophys. J.* **529**, 1031–1033 (2000).
7. Cuntz, M., Saar, S. H. & Musielak, Z. E. *Astrophys. J.* **533**, L151–L154 (2000).
8. Ip, W.-H., Kopp, A. & Hu, J.-H. *Astrophys. J.* **602**, L53–L56 (2004).
9. Lanza, A. F. *Astron. Astrophys.* **487**, 1163–1170 (2008).
10. www.zooniverse.org
11. Shea, M. A., Smart, D. F., McCracken, K. G., Dreschhoff, G. A. M. & Spence, H. E. *Adv. Space Res.* **38**, 232–238 (2006).

Peroxiredoxins are conserved markers of circadian rhythms

Rachel S. Edgar^{1*}, Edward W. Green^{2*}, Yuwei Zhao^{3*}, Gerben van Ooijen⁴, Maria Olmedo⁵, Ximing Qin³, Yao Xu³, Min Pan⁶, Utham K. Valekunja¹, Kevin A. Feeney¹, Elizabeth S. Maywood⁷, Michael H. Hastings⁷, Nitin S. Baliga⁶, Martha Merrow⁵, Andrew J. Millar^{4,8}, Carl H. Johnson³, Charalambos P. Kyriacou², John S. O'Neill¹ & Akhilesh B. Reddy¹

Cellular life emerged ~3.7 billion years ago. With scant exception, terrestrial organisms have evolved under predictable daily cycles owing to the Earth's rotation. The advantage conferred on organisms that anticipate such environmental cycles has driven the evolution of endogenous circadian rhythms that tune internal physiology to external conditions. The molecular phylogeny of mechanisms driving these rhythms has been difficult to dissect because identified clock genes and proteins are not conserved across the domains of life: Bacteria, Archaea and Eukaryota. Here we show that oxidation–reduction cycles of peroxiredoxin proteins constitute a universal marker for circadian rhythms in all domains of life, by characterizing their oscillations in a variety of model organisms. Furthermore, we explore the interconnectivity between these metabolic cycles and transcription–translation feedback loops of the clockwork in each system. Our results suggest an intimate co-evolution of cellular timekeeping with redox homeostatic mechanisms after the Great Oxidation Event ~2.5 billion years ago.

Circadian rhythms are considered to be a feature of almost all living cells. When isolated from external stimuli, organisms exhibit self-sustained cycles in behaviour, physiology and metabolism, with a period of approximately 24 h¹. Circadian clocks afford competitive selective advantages that have been observed experimentally^{2,3}, and disturbance of circadian timing in humans, as seen in rotational shift work and jet lag, carries long-term health costs⁴. For all organisms in which the molecular timing mechanism has been investigated, a common model has arisen, namely a transcription–translation feedback loop (TTFL). TTFL components are not, however, shared between organisms. For example, the cyanobacterial clock is modelled around three proteins: KaiA, B and C. In the fungus *Neurospora crassa*, a loop involving the protein FREQUENCY (FRQ) and the WHITE COLLAR (WC) complex is thought to drive cellular rhythms, whereas the plant TTFL involves elements including TOC1 and CCA1 (refs 1, 5). Furthermore, although some multicellular organisms such as *Drosophila* and humans possess homologous components (for example, the Period proteins), their functions seem to differ between organisms^{6–8}. Therefore, across phylogenetic kingdoms, there are apparently no common 'clock' components, suggesting that daily timekeeping evolved independently within different lineages. The converse, however, could equally be true, and the primary premise of this study was therefore to test the hypothesis that circadian clocks may instead have a common ancestry.

Conservation of peroxiredoxin in circadian systems

Recent studies show that the oxidation state of highly conserved peroxiredoxin (PRX) proteins exhibit circadian oscillations in cells from humans, mice and marine algae^{9,10}, probably reflecting an endogenous rhythm in the generation of reactive oxygen species (ROS)¹⁰. Because

virtually all living organisms possess peroxiredoxins¹¹, we proposed that this marker for circadian rhythms in metabolism may be functionally conserved across all three phylogenetic domains: Archaea, Bacteria and Eukaryota. Peroxiredoxins are peroxidases, the activity of which is dependent on the oxidation of a key 'peroxidatic' cysteine residue (C_p) in the active site, that is absolutely conserved, as are neighbouring proline and threonine/serine residues: conforming to a PXXX(T/S)XXC_p consensus¹². Crucially, the catalytic cysteine can become over- or hyperoxidized (PRX-SO_{2/3}), rendering the peroxiredoxin catalytically inactive, but able to participate in ROS signalling and chaperone activity¹³. Once overoxidized, peroxiredoxin can be recycled by sulphiredoxin.

We previously characterized circadian cycles of peroxiredoxin oxidation using antiserum directed against the oxidized active site, which recognizes both over- (PRX-SO₂) and hyper- (PRX-SO₃) oxidized forms^{9,10,14}. To determine whether this antiserum (raised against an oxidized DFTFVCPTEI peptide) could be used to assay over-/hyper-oxidation in diverse species, we performed several sequence alignments to compare peroxiredoxin protein sequences across a variety of circadian model organisms. This revealed a remarkable degree of conservation across all phylogenetic domains, especially within the active site (Fig. 1a, Supplementary Fig. 1 and Supplementary Tables 1–7). Even when we examined the structure of HyrA, the most distantly related peroxiredoxin orthologue in the archaeon *Halobacterium salinarum* sp. NRC-1, we found that amino acid substitutions would not perturb the geometry of the active site (Fig. 1b). Together, these findings suggested that the same antiserum could be used to probe oxidation rhythms in potentially any organism expressing a peroxiredoxin protein. This was confirmed by gene knockout and peroxide treatment in several representative organisms (Supplementary Figs 3 and 7–9).

¹Department of Clinical Neurosciences, University of Cambridge Metabolic Research Laboratories, NIHR Biomedical Research Centre, Institute of Metabolic Science, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. ²Department of Genetics, University of Leicester, Leicester LE1 7RH, UK. ³Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235-1634, USA. ⁴Synthetic and Systems Biology (SynthSys), Mayfield Road, Edinburgh EH9 3JD, UK. ⁵Department of Molecular Chronobiology, Center for Life Sciences, University of Groningen, 9700 CC Groningen, The Netherlands. ⁶Institute for Systems Biology, 401 Terry Avenue North, Seattle, Washington 98109, USA. ⁷MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. ⁸School of Biological Sciences, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, UK.

*These authors contributed equally to this work.

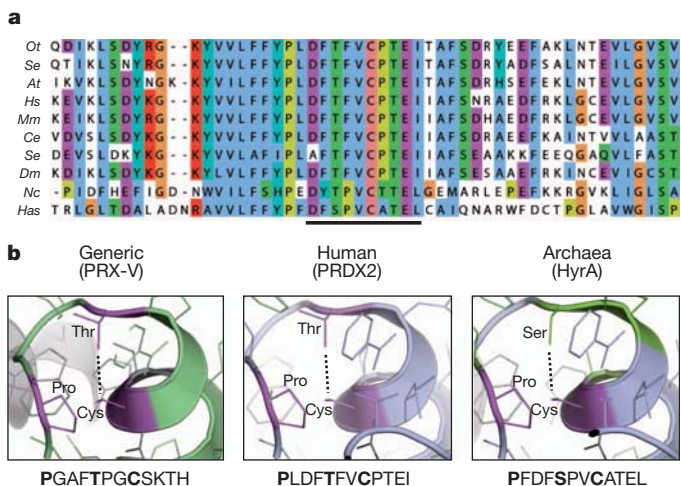


Figure 1 | The peroxiredoxin active site is highly conserved in all domains of life. **a**, Multiple sequence alignment showing peroxiredoxin amino acid sequences. The highly conserved active site is underlined. Representatives shown from Eukaryota (*At*, *A. thaliana*; *Ce*, *Caenorhabditis elegans*; *Dm*, *D. melanogaster*; *Hs*, *Homo sapiens*; *Mm*, *M. musculus*; *Nc*, *N. crassa*; *Ot*, *O. tauri*; *Sc*, *S. cerevisiae*), Bacteria (*Se*, *S. elongatus* sp. PCC7942) and Archaea (*Has*, *H. salinarum* sp. NRC-1). **b**, Critical residues in the active site of 2-Cys peroxiredoxins (in bold) are conserved in all organisms. Structures were derived from human PRX-V (Protein Data Bank (PDB) accession 1HD2)⁴⁷ and human PRDX2 (PDB accession 1QMV)⁴⁸, and modified with PyMOL to show the predicted structure for archaeal peroxiredoxin (HyrA, GenBank accession NP_280562.1).

Peroxiredoxin rhythms in eukaryotes

Using the PRX-SO_{2/3} antiserum, we first examined circadian time courses from a range of eukaryotes under constant conditions (that is, in the absence of external timing cues). In mice, PRX-SO_{2/3} and total PRX1 exhibited a daily cycle in liver tissue and also in the central pacemaker, the suprachiasmatic nuclei (SCN) of the hypothalamus (Fig. 2a). Interestingly, between the two tissues, peroxiredoxin oxidation rhythms were in distinct, and different, phase relationships with respect to total PRX1 and BMAL1 protein, suggesting a difference between brain and peripheral tissue (Fig. 2a), as observed previously for other clock components¹⁵.

To extend these findings beyond vertebrates, we examined peroxiredoxin rhythms in the fruitfly *Drosophila melanogaster*. We pooled whole heads from insects maintained in constant darkness over two circadian cycles after they had been stably entrained to 12 h light, 12 h dark cycles. Again, circadian oscillations in PRX-SO_{2/3} immunoreactivity were observed, as well as in the clock protein Timeless (TIM) (Fig. 2b). Similarly, seedlings from the plant *Arabidopsis thaliana* exhibited robust PRX-SO_{2/3} oscillations in free-running conditions of constant light, which were also seen in the filamentous fungus *Neurospora crassa*—another well-characterized clock model system (Fig. 2c and Supplementary Fig. 2). Therefore, just as in tissue from ‘complex’ vertebrates, this range of ‘simpler’ eukaryotic systems had robust peroxiredoxin oxidation cycles, with peaks tending to occur around anticipated dawn.

Peroxiredoxin rhythms in prokaryotes

Having observed ~24 h peroxiredoxin oxidation rhythms in organisms with nucleated cells, we next sought to examine representative prokaryotes from each major domain—Bacteria and Archaea.

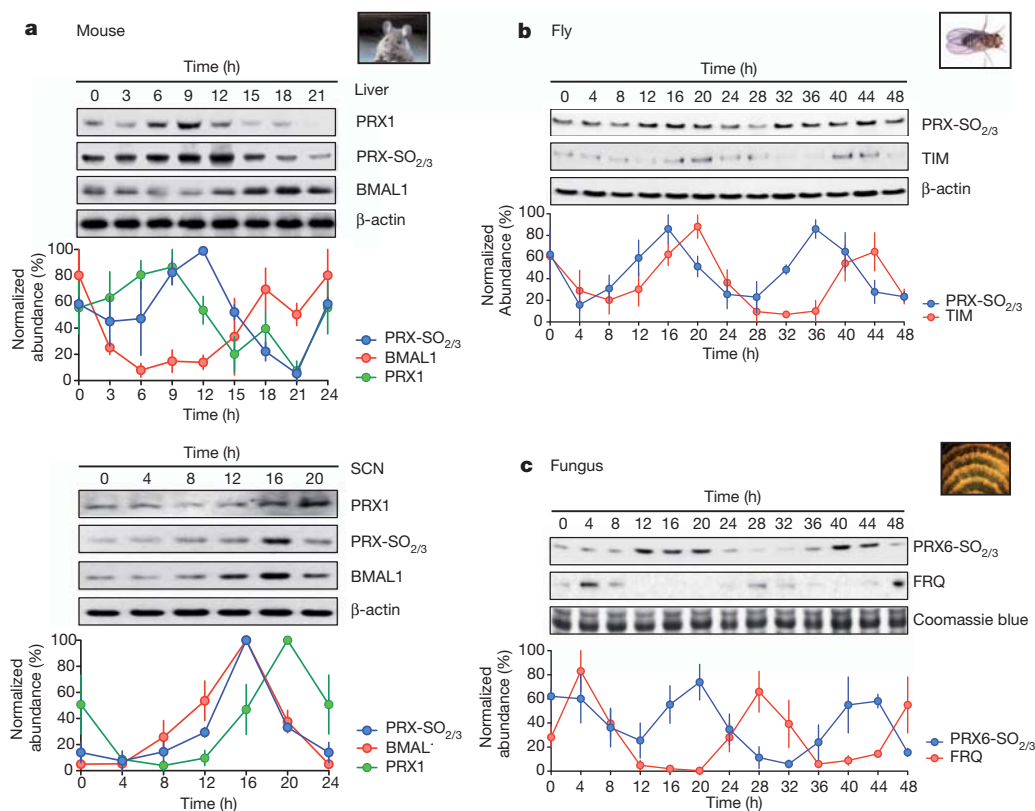


Figure 2 | Peroxiredoxin oxidation cycles are conserved in eukaryotic models of the circadian clock. **a–c**, Representative immunoblots probed for oxidized/hyperoxidized 2-Cys peroxiredoxin (PRX-SO_{2/3} or PRX6-SO_{2/3}) are shown for mouse (*Mus musculus*; **a**), fruitfly (*D. melanogaster*; **b**), and fungus (*N. crassa*; **c**). For each model system, the organism was sampled under free-running conditions. Loading controls show either β -actin immunoblots or

Coomassie blue-stained gels loaded with identical samples used for immunoblotting. Immunoblot quantification by densitometry is shown below each panel (mean \pm s.e.m.) for $n = 3$ biological replicates. See Supplementary Fig. 2 for plant rhythms, and Supplementary Table 9 for cycle period estimates (by harmonic regression) and detailed statistics (by ANOVA).

For bacteria, we used the best characterized prokaryotic clock model system, *Synechococcus elongatus* sp. PCC7942 (ref. 16). The major proteins involved in the cyanobacterial clockwork are encoded by the *KaiABC* cluster from which, remarkably, circadian oscillations of phosphorylation can be reconstituted *in vitro*¹⁷. However, because all three Kai proteins are expressed together in only a small number of bacterial species, and in no known archaea¹⁸, this system cannot represent a general prokaryotic clock mechanism. We postulated that, regardless of the timekeeping mechanism, the peroxiredoxin oxidation cycles we observe reflect an absolutely conserved rhythmic cellular output. We tested this by assaying PRX-SO_{2/3} under free-running conditions (constant light) for 48 h, and observed cycles of cyanobacterial peroxiredoxin oxidation, peaking later than phosphorylated KaiC (Fig. 3a).

Furthermore, we extended our studies to the third phylogenetic domain, Archaea, assaying rhythms of peroxiredoxin oxidation in *H. salinarum* sp. NRC-1. Although no clock mechanisms have been identified for any archaeon, diurnal transcriptional rhythms were recently observed in *H. salinarum*¹⁹, and it thus represents an ideal platform to test whether peroxiredoxin oxidation constitutes a universal marker for cellular rhythms. After entraining the archaea for three cycles in 12 h light, 12 h dark cycles, we placed them into constant light at constant temperature. We observed robust, high amplitude circadian oscillations of PRX-SO_{2/3} for three cycles (Fig. 3b). Together, our findings in evolutionarily diverse prokaryotes provide compelling evidence that rhythmic peroxiredoxin oxidation is a conserved circadian marker across phylogenetic domains.

Relations between peroxiredoxin cycles and TTFLs

In all circadian model systems, the proposed clock mechanism revolves around a TTFL^{5,20,21}. How the metabolic rhythms observable

through peroxiredoxin oxidation relate to, and interact with, the known transcriptional clockwork in different organisms is unclear. Again, we used several model organisms, with available 'clock' mutants, to address this question.

The *Drosophila* transcriptional clockwork is structurally similar to the mammalian TTFL, although there are some important differences²². In flies, Clock and Cycle (orthologous to mammalian BMAL1, also known as ARNTL) comprise the positive limb, driving oscillatory expression of Period (PER) and Timeless (TIM). PER and TIM negatively regulate their own expression, closing the loop^{5,6}. This circuit can be disrupted by many mutations²³. Two mutants, *per⁰¹* and *Clk^{rk}*, are behaviourally arrhythmic, and show non-cycling expression of circadian components, including PER and TIM^{24,25}. To examine peroxiredoxin oxidation patterns in these mutants, they were entrained as described above for wild-type (Canton-S) flies. We observed two circadian cycles of PRX-SO_{2/3} oscillation, with an altered circadian phase relative to wild type. This indicates the presence of an underlying capacity for circadian timing in both mutant strains, which was clearly perturbed by the absence of functional transcriptional feedback circuitry (Fig. 4a).

To establish the wider relevance of these findings, we also examined similar mutants in the fungus *Neurospora crassa*. The frequency (*frq*) locus encodes a critical element in the TTFL of *Neurospora*, in addition to the Per-Arnt-Sim (PAS)-containing WC transcription factors²⁶. In the long-period *frq⁷* mutant²⁷, peroxiredoxin oxidation rhythms showed a similarly lengthened period, with an altered phase relative to rhythms in FRQ protein abundance (Fig. 4b). Deletion of the *frq* locus characterizes the *frq¹⁰* strain, and measurable markers of clock output, such as its spore-forming ('conidiation') rhythm, are profoundly perturbed in these fungi, although apparently stochastic oscillations can re-emerge under various growth conditions^{27–29}. Circadian rhythms of peroxiredoxin oxidation were, however, clearly seen in *frq¹⁰* mutants sampled in constant darkness (Fig. 4b), with a delayed phase relative to wild-type (*bd*) fungi. This illustrates that peroxiredoxin rhythms represent an alternative readout for an oscillator that persists in the absence of a FRQ-dependent clock.

We next examined the phenotypes of mutant circadian transcriptional regulators in photosynthetic eukaryotes and prokaryotes. The transcriptional clockwork of the plant *Arabidopsis thaliana* and the alga *Ostreococcus tauri* are very similar and rely on circadian oscillation of TOC1. Accordingly, overexpressing *TOC1* in either species disrupts transcriptional rhythms^{30,31}. In such strains, under constant light, we observed persistent oscillations of peroxiredoxin oxidation, albeit with altered amplitude and phase relative to controls (Supplementary Figs 2 and 3). Furthermore, in cyanobacteria we assayed peroxiredoxin oxidation in the arrhythmic *KaiA* deletion strain, *AMC702* (ref. 32). Notably, an approximately 24-h rhythm of peroxiredoxin oxidation persisted despite a functional Kai-based oscillator being absent, again in an altered phase relative to wild type (Fig. 5a). Taken together, these observations indicate that metabolic rhythms remain closely aligned to transcriptional feedback mechanisms when those mechanisms are present. Crucially, however, metabolic rhythms persist even when cycling clock gene transcription is abolished (summarized in Supplementary Table 9).

Having determined the TTFL influence on peroxiredoxin oxidation rhythms, we reciprocally tested whether rhythmic peroxiredoxin oxidation is required for timekeeping, using TTFL components as markers of the clockwork. We assayed reporter bioluminescence and delayed fluorescence in mutant *S. elongatus* and *A. thaliana* lines, respectively, that were deficient in 2-CysPRX (*Synechococcus* Δ 2-CysPRX, GenBank accession AAP49028; *Arabidopsis* double mutant: Δ 2-CysPRXA Δ 2-CysPRXB, GenBank accessions NM_111995 and NM_120712)³³. In these mutants, circadian rhythms persisted with wild-type period, albeit significantly perturbed in either phase or amplitude, relative to controls (Fig. 5b and Supplementary Fig. 9). This suggests that peroxiredoxins are not required for oscillator

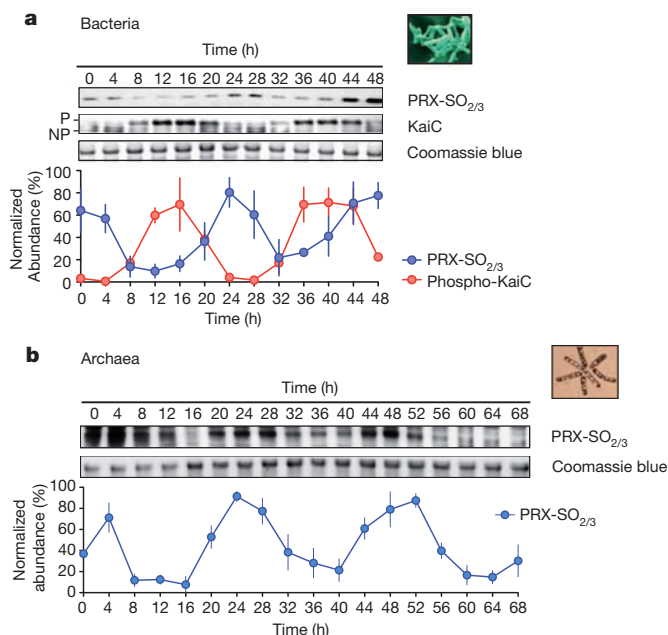


Figure 3 | Peroxiredoxin oxidation cycles are conserved in prokaryotic models of the circadian clock. **a, b**, Representative immunoblots probed for oxidized/hyperoxidized 2-Cys peroxiredoxin (PRX-SO_{2/3}) are shown for bacteria (*S. elongatus* sp. PCC7942; **a**), and archaea (*H. salinarum* sp. NRC-1; **b**). Before sampling under free-running conditions (constant light), cyanobacteria were synchronized with a 12 h dark pulse, whereas archaea were stably entrained to 12 h light, 12 h dark cycles. Loading controls show Coomassie blue-stained gels loaded with identical samples used for immunoblotting. Immunoblot quantification by densitometry is shown below each panel (mean \pm s.e.m.) for $n = 3$ biological replicates. See Supplementary Table 9 for cycle period estimates and detailed statistics. P, phosphorylated KaiC; NP, non-phosphorylated KaiC.

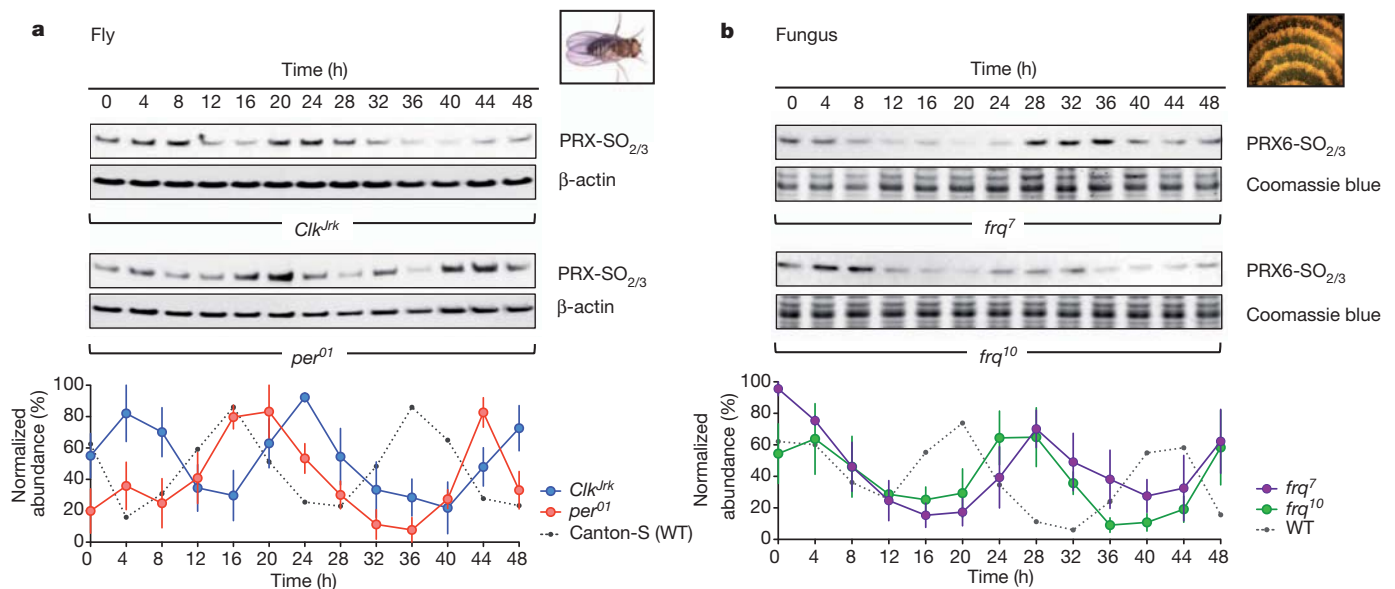


Figure 4 | Peroxiredoxin oxidation cycles in circadian clock mutants. **a, b,** Representative immunoblots probed for oxidized/hyperoxidized peroxiredoxin (PRX-SO_{2/3} or PRX6-SO_{2/3}) are shown for fruitfly (*D. melanogaster*; **a**) and fungus (*N. crassa*; **b**). For each model system, organisms were sampled under free-running conditions (constant darkness). Loading controls show either β-actin immunoblots or Coomassie blue-stained gels

function in systems that possess an alternative timing mechanism, namely a TTFL. On the other hand, our findings in TTFL mutants reveal that cellular components that are required for rhythmic outputs are not essential to rhythms in redox metabolism. Given the information that we have about the above model organisms, used commonly to study clock biology, we suggest that both peroxiredoxin and TTFL components of the circadian system are important, but potentially individually dispensable for circadian rhythms at the cellular level. Moreover, the phenotypes of these mutants suggest that the cellular ROS balance is important for robust clock function, as was described recently in *Neurospora*³⁴.

Implications for clock evolution

We have observed ~24 h cycles of peroxiredoxin oxidation–reduction in all domains of life and consequently, the possibility that cellular rhythms share a common molecular origin seems increasingly plausible. Because the cellular role of peroxiredoxins principally involves the removal of toxic metabolic by-products (that is, ROS), we proposed that the ability to survive cycles of oxidative stress may have contributed a selective advantage from the beginnings of aerobic life.

Approximately 2.5 billion years ago, photosynthetic bacteria acquired the capacity for photo-dissociation of water, leading to the geologically rapid accumulation of molecular oxygen during the Great Oxidation Event (GOE), when anaerobic life underwent a catastrophic decline³⁵. Evidently, organisms that survived the transition to an aerobic environment were those that respired and/or evolved oxygen. As electron transport chains involving oxygen inevitably produce toxic superoxide anions as by-products³⁶, during the GOE, successful organisms had to acquire ROS removal systems or were relegated to anaerobic niches³⁵. Superoxide dismutase, which converts superoxide to hydrogen peroxide, is ubiquitous and, like peroxiredoxin (the major cellular H₂O₂ ‘sink’), is estimated to have arisen around the time of the GOE³⁷, the same era during which the most ancient known clock mechanism (the Kai oscillator) evolved. Importantly, we note that (1) during the GOE, rhythms of O₂ production/consumption and ROS generation would have been driven by the solar cycle, as they are today^{38–40}; (2) metabolic/oxidation rhythms

loaded with identical samples used for immunoblotting. Immunoblot quantification by densitometry is shown below each panel (mean ± s.e.m.) for *n* = 3 biological replicates. See Supplementary Table 9 for cycle period estimates (by harmonic regression) and detailed statistics (by ANOVA), as well as Supplementary Fig. 10 for TIM and FRQ immunoblots for fruitfly and fungus, respectively. WT, wild type.

seem to be present in every organism with a circadian clock, all of which are aerobes, and these rhythms persist in the absence of transcriptional cycles; and (3) circadian timekeeping confers a selective advantage when it facilitates anticipation of environmental change (noxious or otherwise).

We believe that the most reasonable interpretation is that cellular metabolism in the most competitive early aerobes adapted to confer anticipation of, and resonate with, environmental cycles in energy supply and oxidative stress. We presume that the echoes of this ancient evolutionary adaptation are revealed by the conserved peroxiredoxin oxidation cycles we now observe in disparate organisms, indicating that during the past 2.5 billion years, ROS and metabolic pathways must have co-evolved with the cellular clockwork, and are probably interlinked. On top of this, further cellular mechanisms seem to have been incorporated, when advantageous, as they arose over time. For example, in eukaryotes, the timekeeping contribution made by certain post-translational mechanisms (such as casein kinase) is conserved across several disparate organisms, whereas clock gene transcription factors are widely divergent, having evolved and been introduced more recently (see Fig. 6).

If there was considerable pressure for the co-evolution of metabolic/ROS pathways with cellular timekeeping systems, then evidence for this should exist in the phylogenetic trees of their component mechanisms. To substantiate this we used the Mirrortree algorithm⁴¹ to assess the degree of co-evolution between the 2-Cys peroxiredoxin family, representing metabolism/ROS pathways, with the most ancient characterized clock mechanism: the three cyanobacterial Kai proteins. Because the Kai proteins are found exclusively in prokaryotes, we focused on these for our analysis¹⁸. All three components of the cyanobacterial oscillator seem to have co-evolved with 2-Cys peroxiredoxins, as shown by the strong correlation between the distances of respective proteins within each phylogenetic tree (KaiA, *r* = 0.784; KaiB, *r* = 0.883; KaiC, *r* = 0.865; *P* < 1 × 10^{−6} for all) (Fig. 5c and Supplementary Fig. 4). Notably, when evolution of KaiC (the most ancient member) was compared with other absolutely conserved protein families, the three highest correlations observed were for the other two clock components (KaiA and KaiB) and for

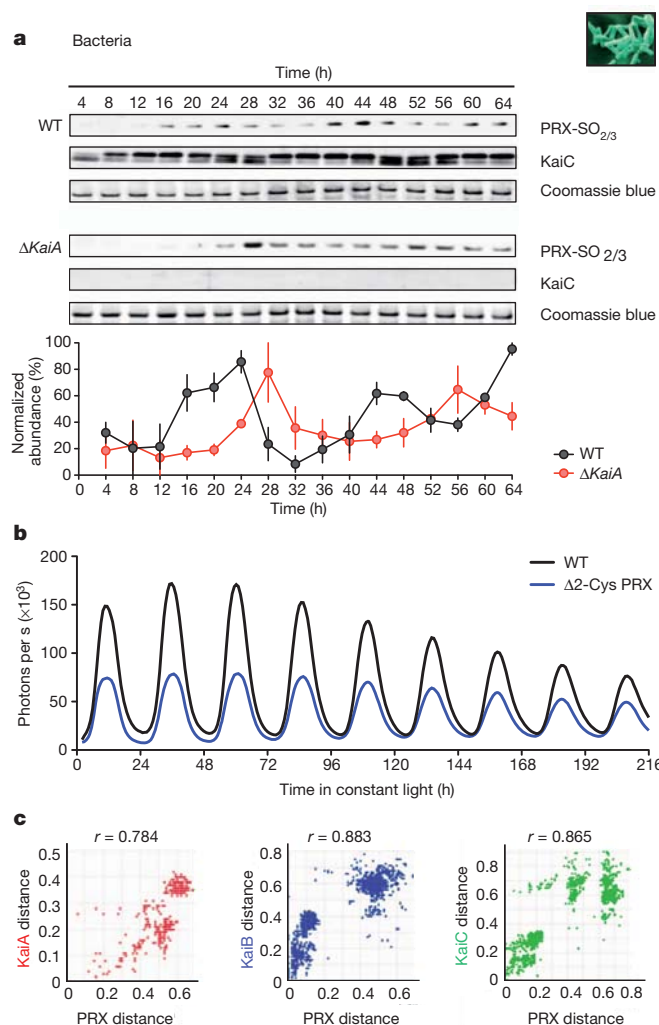


Figure 5 | Relationships between peroxiredoxins and the cyanobacterial Kai-based oscillator. **a**, Representative immunoblots for oxidized/hyperoxidized 2-Cys peroxiredoxin (PRX-SO_{2/3}) are shown for wild-type (strain AMC149) and KaiA deletion mutant (ΔKaiA; strain AMC702) cyanobacteria. Immunoblot quantification by densitometry is shown below each panel (mean \pm s.e.m.), $n = 3$ biological replicates. See Supplementary Table 9 for detailed analysis. **b**, Bioluminescence traces for cyanobacteria cultures of wild-type or 2-Cys peroxiredoxin knockout (Δ2-Cys PRX) strains, as reported by *psbAip::luxAB*⁴⁹. **c**, Co-evolution of cyanobacteria oscillator components and peroxiredoxin proteins. Interspecies plots correlate the interprotein distances between Kai proteins and 2-Cys peroxiredoxin^{41,50}. r denotes the correlation coefficient ($P < 1 \times 10^{-6}$). See Supplementary Table 8 and Supplementary Figs 4–6 for details.

peroxiredoxin (Supplementary Fig. 5 and Supplementary Table 8). This suggests that similarities in the evolutionary profiles of these cellular mechanisms go beyond those that would be expected simply based on the time since a common ancestry, because even highly conserved proteins had considerably inferior correlations to peroxiredoxin (Supplementary Fig. 6 and Supplementary Table 8)⁴². Given that Kai proteins are not found in eukaryotic systems, but metabolic rhythms such as those observed in peroxiredoxin oxidation are, our results suggest that metabolic rhythms are at least as ancient as, and pre-date most, phylum-specific timekeeping mechanisms, but are intrinsically integrated with them in modern organisms.

Concluding remarks

It has long been recognized that oxygen-sensing PAS-domain proteins are involved in the clockwork of many eukaryotes, but the rationale behind this has remained elusive^{23,43–45}. In light of our current findings,

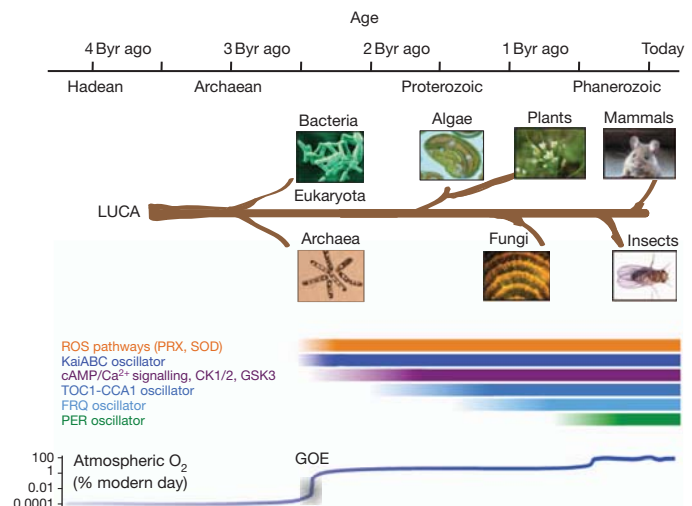


Figure 6 | Phylogenetic origins of circadian oscillatory systems. A timeline is shown at the top of the schematic, with the geological era illustrated. A schematic phylogenetic tree shows the origins of each organism studied, stemming from the last universal common ancestor (LUCA). The putative epoch over which each oscillator system has existed is illustrated by the labelled bars. CK1/2, casein kinase 1 or 2; GSK3, glycogen synthase kinase 3; SOD, superoxide dismutase.

we speculate that sensing and responding to oxidative cycles in cellular environments could have driven the evolution of circadian rhythms, and maintained the intrinsic link between clocks and metabolism (Fig. 6). A direct prediction therefore, is that organisms that lack ROS detoxification systems will not have circadian rhythms. At least one such class of organism exists on Earth, an example being the hyperthermophilic archaea *Methanopyri* (NCBI taxonomy accession 183988). Given its distinct anoxic environmental niche and methanogenic metabolism⁴⁶, there may never have been a selective evolutionary pressure to develop circadian timekeeping as we know it.

METHODS SUMMARY

Organisms and strains. *A. thaliana*, *D. melanogaster*, *H. salinarum* sp. NRC-1, *M. musculus*, *N. crassa*, *O. tauri* and *S. elongatus* were bred, grown or cultured in appropriate conditions, synchronized by specific methods normally used in each organisms, and then sampled under constant conditions of either darkness or light, depending on the organism.

Gel electrophoresis and western blotting. Samples were lysed in LDS buffer and heated to 70 °C for 10 min with constant shaking (600 r.p.m.) in a thermomixer. Electrophoresis was performed using pre-fabricated 4–12% Bis-Tris gradient gels, using a non-reducing MES SDS buffer system, allowing characterization of proteins between 10 and 260 kDa. Immunoblotting was performed after protein transfer to nitrocellulose membranes. After blocking, membranes were incubated in antibody diluted in blocking buffer (0.5% milk in BSA) overnight at 4 °C. The next day, membranes were washed and bands visualized with chemiluminescence detection.

Phylogenetic analyses. We used the Mirrortree online server, using input FASTA sequences for human PRDX2 (GenBank accession CAG46588.1) and compared this serially with *S. elongatus* sp. PCC7942 proteins: KaiA (GenBank accession AAM82684.1), KaiB (AAM82685.1) and KaiC (AAM82686.1). Similar analyses were performed for KaiC comparisons with other conserved bacterial proteins. Interspecies plots were generated, which contain a simplified representation of the correlation between the interprotein distances in phylogenetic trees for each protein being compared.

Image and statistical analysis. Coomassie-stained gel images were obtained using a Licor Odyssey system, and immunoblot films were scanned using a flat-bed scanner. Densitometric quantification of images was performed using NIH ImageJ software. Signal was normalized against the respective loading control for each replicate at each time point for grouped data. Sine curve fitting was performed using Circwave software, using a harmonic regression method, and analysis of variance (ANOVA) was also performed as an independent measure of temporal variation. Statistical comparisons between Mirrortree correlation coefficients were performed as detailed above.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 July 2011; accepted 26 March 2012.

Published online 16 May 2012.

- Dunlap, J. C. Molecular bases for circadian clocks. *Cell* **96**, 271–290 (1999).
- Woelfle, M. A., Ouyang, Y., Phanvijitsiri, K. & Johnson, C. H. The adaptive value of circadian clocks; an experimental assessment in cyanobacteria. *Curr. Biol.* **14**, 1481–1486 (2004).
- Dodd, A. N. *et al.* Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* **309**, 630–633 (2005).
- Barger, L. K., Lockley, S. W., Rajaratnam, S. M. & Landrigan, C. P. Neurobehavioral, health, and safety consequences associated with shift work in safety-sensitive professions. *Curr. Neurol. Neurosci. Rep.* **9**, 155–164 (2009).
- Wijnen, H. & Young, M. W. Interplay of circadian clocks and metabolic rhythms. *Annu. Rev. Genet.* **40**, 409–448 (2006).
- Allada, R., Emery, P., Takahashi, J. S. & Rosbash, M. Stopping time: the genetics of fly and mouse circadian clocks. *Annu. Rev. Neurosci.* **24**, 1091–1119 (2001).
- Rosbash, M. The implications of multiple circadian clock origins. *PLoS Biol.* **7**, e62 (2009).
- Zheng, X. & Sehgal, A. Probing the relative importance of molecular oscillations in the circadian clock. *Genetics* **178**, 1147–1155 (2008).
- O'Neill, J. S. *et al.* Circadian rhythms persist without transcription in a eukaryote. *Nature* **469**, 554–558 (2011).
- O'Neill, J. S. & Reddy, A. B. Circadian clocks in human red blood cells. *Nature* **469**, 498–503 (2011).
- Hall, A., Karplus, P. A. & Poole, L. B. Typical 2-Cys peroxiredoxins—structures, mechanisms and functions. *FEBS J.* **276**, 2469–2477 (2009).
- Wood, Z. A., Poole, L. B. & Karplus, P. A. Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling. *Science* **300**, 650–653 (2003).
- Barranco-Medina, S., Lazaro, J. J. & Dietz, K. J. The oligomeric conformation of peroxiredoxins links redox state to function. *FEBS Lett.* **583**, 1809–1816 (2009).
- Woo, H. A. *et al.* Reversible oxidation of the active site cysteine of peroxiredoxins to cysteine sulfinic acid. Immunoblot detection with antibodies specific for the hyperoxidized cysteine-containing sequence. *J. Biol. Chem.* **278**, 47361–47364 (2003).
- Lopez-Molina, L., Conquet, F., Dubois-Dauphin, M. & Schibler, U. The DBP gene is expressed according to a circadian rhythm in the suprachiasmatic nucleus and influences circadian behavior. *EMBO J.* **16**, 6762–6771 (1997).
- Johnson, C. H., Mori, T. & Xu, Y. A cyanobacterial circadian clockwork. *Curr. Biol.* **18**, R816–R825 (2008).
- Nakajima, M. *et al.* Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*. *Science* **308**, 414–415 (2005).
- Dvornyk, V., Vinogradova, O. & Nevo, E. Origin and evolution of circadian clock genes in prokaryotes. *Proc. Natl Acad. Sci. USA* **100**, 2495–2500 (2003).
- Whitehead, K., Pan, M., Masumura, K., Bonneau, R. & Baliga, N. S. Diurnally entrained anticipatory behavior in archaea. *PLoS ONE* **4**, e5485 (2009).
- Lakin-Thomas, P. L. Transcriptional feedback oscillators: maybe, maybe not. *J. Biol. Rhythms* **21**, 83–92 (2006).
- Reddy, A. B. & O'Neill, J. S. Healthy clocks, healthy body, healthy mind. *Trends Cell Biol.* **20**, 36–44 (2009).
- Young, M. W. & Kay, S. A. Time zones: a comparative genetics of circadian clocks. *Nature Rev. Genet.* **2**, 702–715 (2001).
- Young, M. W. The molecular control of circadian behavioral rhythms and their entrainment in *Drosophila*. *Annu. Rev. Biochem.* **67**, 135–152 (1998).
- Allada, R., White, N. E., So, W. V., Hall, J. C. & Rosbash, M. A mutant *Drosophila* homolog of mammalian clock disrupts circadian rhythms and transcription of *period* and *timeless*. *Cell* **93**, 791–804 (1998).
- Hardin, P. E. The circadian timekeeping system of *Drosophila*. *Curr. Biol.* **15**, R714–R722 (2005).
- Dunlap, J. C. & Loros, J. J. How fungi keep time: circadian system in *Neurospora* and other fungi. *Curr. Opin. Microbiol.* **9**, 579–587 (2006).
- Aronson, B. D., Johnson, K. A., Loros, J. J. & Dunlap, J. C. Negative feedback defining a circadian clock: autoregulation of the clock gene frequency. *Science* **263**, 1578–1584 (1994).
- Granshaw, T., Tsukamoto, M. & Brody, S. Circadian rhythms in *Neurospora crassa*: farnesol or geraniol allow expression of rhythmicity in the otherwise arrhythmic strains *frq¹⁰*, *wc-1*, and *wc-2*. *J. Biol. Rhythms* **18**, 287–296 (2003).
- Lakin-Thomas, P. L. & Brody, S. Circadian rhythms in microorganisms: new complexities. *Annu. Rev. Microbiol.* **58**, 489–519 (2004).
- Corellou, F. *et al.* Clocks in the green lineage: comparative functional analysis of the circadian architecture of the picoeukaryote *ostreococcus*. *Plant Cell* **21**, 3436–3449 (2009).
- Mas, P., Alabadi, D., Yanovsky, M. J., Oyama, T. & Kay, S. A. Dual role of TOC1 in the control of circadian and photomorphogenic responses in *Arabidopsis*. *Plant Cell* **15**, 223–236 (2003).
- Ditty, J. L., Canales, S. R., Anderson, B. E., Williams, S. B. & Golden, S. S. Stability of the *Synechococcus elongatus* PCC 7942 circadian clock under directed anti-phase expression of the kai genes. *Microbiology* **151**, 2605–2613 (2005).
- Pulido, P. *et al.* Functional analysis of the pathways for 2-Cys peroxiredoxin reduction in *Arabidopsis thaliana* chloroplasts. *J. Exp. Bot.* **61**, 4043–4054 (2010).
- Yoshida, Y., Iigusa, H., Wang, N. & Hasunuma, K. Cross-talk between the cellular redox state and the circadian system in *Neurospora*. *PLoS ONE* **6**, e28227 (2011).
- Wang, M. *et al.* A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. *Mol. Biol. Evol.* **28**, 567–582 (2011).
- Nathan, C. & Ding, A. SnapShot: reactive oxygen intermediates (ROI). *Cell* **140**, 951–951.e2 (2010).
- Zelko, I. N., Mariani, T. J. & Folz, R. J. Superoxide dismutase multigene family: a comparison of the CuZn-SOD (SOD1), Mn-SOD (SOD2), and EC-SOD (SOD3) gene structures, evolution, and expression. *Free Radic. Biol. Med.* **33**, 337–349 (2002).
- Mulholland, P. J., Houser, J. N. & Maloney, K. O. Stream diurnal dissolved oxygen profiles as indicators of in-stream metabolism and disturbance effects: Fort Benning as a case study. *Ecol. Indic.* **5**, 243–252 (2005).
- Venkateswaran, J. J., Wassenaar, L. I. & Schiff, S. L. Dynamics of dissolved oxygen isotopic ratios: a transient model to quantify primary production, community respiration, and air-water exchange in aquatic ecosystems. *Oecologia* **153**, 385–398 (2007).
- Bamforth, S. S. Diurnal changes in shallow aquatic habitats. *Limnol. Oceanogr.* **7**, 348–353 (1962).
- Ochoa, D. & Pazos, F. Studying the co-evolution of protein families with the Mirrortree web server. *Bioinformatics* **26**, 1370–1371 (2010).
- Peixoto, A. A., Campesan, S., Costa, R. & Kyriacou, C. P. Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. *Mol. Biol. Evol.* **10**, 127–139 (1993).
- McIntosh, B. E., Hogenesch, J. B. & Bradfield, C. A. Mammalian Per-Arnt-Sim proteins in environmental adaptation. *Annu. Rev. Physiol.* **72**, 625–645 (2010).
- Rutter, J., Reick, M. & McKnight, S. L. Metabolism and the control of circadian rhythms. *Annu. Rev. Biochem.* **71**, 307–331 (2002).
- Rutter, J., Reick, M., Wu, L. C. & McKnight, S. L. Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. *Science* **293**, 510–514 (2001).
- Shima, S., Thauer, R. K. & Ermler, U. Hyperthermophilic and salt-dependent formyltransferase from *Methanopyrus kandleri*. *Biochem. Soc. Trans.* **32**, 269–272 (2004).
- Declercq, J. P. *et al.* Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 Å resolution. *J. Mol. Biol.* **311**, 751–759 (2001).
- Schröder, E. *et al.* Crystal structure of dimeric 2-Cys peroxiredoxin from human erythrocytes at 1.7 Å resolution. *Structure* **8**, 605–615 (2000).
- Xu, Y., Mori, T. & Johnson, C. H. Cyanobacterial circadian clockwork: roles of KaiA, KaiB and the *kaiBC* promoter in regulating KaiC. *EMBO J.* **22**, 2117–2126 (2003).
- Pazos, F. & Valencia, A. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng.* **14**, 609–614 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was primarily supported by the Wellcome Trust (083643/Z/07/Z and 093734/Z/10/Z), the European Research Council (ERC Starting Grant No. 281348, MetaCLOCK), and EMBO Young Investigators Programme, as well as the Medical Research Council Centre for Obesity and Related Metabolic Disorders (MRC CORD), and the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre. C.P.K. and M.H.H. acknowledge European Commission grant EUCLOCK (no. 018741) and Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/C006941/1. SynthSys is funded by BBSRC and Engineering and Physical Sciences Research Council (EPSRC) award BB/D019621 to A.J.M. and others. N.S.B. was supported by ENIGMA, US Department of Energy, under contract no. DE-AC02-05CH11231, and by a grant from the National Institutes of Health (NIH; P50GM076547). C.H.J. was supported by the NIH (R01GM088595, R01GM067152 and R21HL102492). M.M. was supported by the Netherlands Organisation for Scientific Research (NWO; Dutch Science Foundation VICI award and Open Programma) and the University of Groningen (Rosalind Franklin Fellowship Program). We thank M. Jain, G. O'Neill and J. Chambers for discussion about the manuscript, and S. G. Rhee, F. Rouyer and R. Stanewsky for the gifts of antisera.

Author Contributions A.B.R. and J.S.O. conceived and designed the experiments, and wrote the manuscript. R.S.E., E.W.G., G.v.O., M.O., X.Q., Y.X., Y.Z., M.P., U.K.V., K.A.F. and E.S.M. performed experiments. M.H.H., N.S.B., C.H.J., M.M., A.J.M. and C.P.K. provided reagents. R.S.E., E.W.G., G.v.O., M.O. and Y.Z. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.B.R. (areddy@cantab.net) or J.S.O. (js22@medschl.cam.ac.uk).

METHODS

***Arabidopsis thaliana*.** Surface-sterilized *A. thaliana* seeds (Ws) were plated on solid medium (1.2% agar plus 0.5× Murashige & Skoog medium (Duchefa Biochemie), pH 5.8), vernalized at 4 °C for 4 days, and grown for 7 days under 12 h light, 12 h dark cycles under cool-white fluorescent tubes (70–100 $\mu\text{Em}^{-2}\text{s}^{-1}$) at 20 °C before transfer to constant light conditions at Zeitgeber time (ZT) 0. Plantlets were sampled every 4 h for 3 days by snap-freezing 15 seedlings per sample in liquid N₂. Tissue was crushed in a Tissue Lyser (Eppendorf) using a 0.32-cm chrome ball (Spheric-Trafalgar), and tissue was thawed in extraction buffer (8 M urea, 300 mM NaCl, 100 mM Tris, pH 7.5, 10 mM EDTA, 4% poly(vinylpyrrolidone), 1% NP40 and 2× complete protease inhibitors (Roche)) and incubated on ice for 15 min, vortexing every 2 min before centrifugation at 16,000g for 10 min. Supernatants were loaded onto gels after the addition of SDS loading buffer (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris-HCl, ~pH 6.8) and heating to 100 °C for 5 min. Equal protein loading was confirmed by gel electrophoresis and Coomassie staining of gels loaded with equal volumes of lysate from each time point in each replicate set. Delayed fluorescence was performed as reported previously⁵¹.

***Drosophila melanogaster*.** Canton-S (CS) wild-type flies and congenic *per⁰¹* and *Clk^{rk}* mutants were raised on standard medium at 25 °C in 12 h light, 12 h dark cycles, and their behavioural phenotypes were validated using DAM5 activity monitors (TriKinetics)⁵². Adult male flies were entrained for a further 3 days in light boxes at 25 °C before transfer into constant darkness at 25 °C. Flies were collected over two circadian cycles every 4 h, and their heads were dissociated on dry ice in the light and frozen at –80 °C until use. Protein was extracted from 20 fly heads per time point in CHAPS/urea buffer (8 M urea, 4% (w/w) CHAPS, 5 mM magnesium acetate, 10 mM Tris, pH 8.0) with complete protease inhibitors (Roche)⁵³. Samples were ground in a 1.5-ml microfuge tube with a miniature pestle. Protein-adjusted supernatants were loaded onto gels after adding appropriate volumes of 4× LDS loading buffer (Invitrogen) to bring the final concentration to 2×, and then heating to 70 °C for 10 min. Equal protein loading was confirmed by gel electrophoresis and Coomassie staining of gels loaded with equal volumes of lysate from every time point in each replicate set.

***Halobacterium salinarum* sp. NRC-1.** Wild-type *H. salinarum* sp. NRC-1 was cultured from a single colony in complete medium at 37 °C with shaking at 125 r.p.m. (Innova Waterbath, New Brunswick Scientific)⁵⁴. We performed experiments using three different starting colonies or biological replicates. Cells were incubated under entrainment conditions for 5 days (12 h light, 12 h dark cycle; daylight was simulated with full spectrum light at 150 $\mu\text{Em}^{-2}\text{s}^{-1}$) before sampling. After entrainment, samples (2 ml) were collected every 4 h for 3 days in continuous light, at constant cell density. The cell density was maintained by replacing a fixed volume of the culture with equivalent fresh complete medium (30–80 ml) every 4 h⁵⁵. Absorbance at 600 nm ($A_{600\text{ nm}}$) was recorded at each time point, before dilution, to determine how much medium to add, and also to calculate cell numbers for lysate preparation. As previously described, comparative analysis with a similarly processed control culture discounted any unaccounted perturbations that could have been introduced by this periodic dilution⁵⁴. Cell pellets were collected by centrifugation at 1,600g for 2 min, decanted, flash-frozen in liquid N₂ and stored at –80 °C until protein extraction. Cell pellets were lysed with 2× denaturing LDS sample buffer (Invitrogen) without reducing agent. Total protein was normalized by the addition of 2× LDS sample buffer on the basis of cell numbers in each sample (determined from $A_{600\text{ nm}}$ measurements). Lysates were then passed through a 27-gauge, 1.27-cm needle by syringing (to reduce viscosity of the samples), before heating to 70 °C for 10 min to fully denature proteins before loading on gels. Equal protein loading was confirmed by gel electrophoresis and Coomassie staining of gels loaded with equal volumes of lysate from every time point in each replicate set.

***Mus musculus*.** All animal experimentation was licensed by the Home Office under the Animals (Scientific Procedures) Act 1986, with Local Ethical Review by the Medical Research Council and the University of Cambridge, UK.

For free-run liver experiments. Liver tissue was collected from four adult male C57BL/6 mice once every 3 h on the second cycle after transfer from 12 h light, 12 h dim red light ('dark') to 12 h dim red light, 12 h dim red light (light, 220 $\mu\text{W cm}^{-2}$; dim red light, <5 $\mu\text{W cm}^{-2}$) and immediately frozen and then stored at –80 °C before use. Before sampling, animals were stably entrained to a 12 h light, 12 h dim red light cycle for 3–4 weeks. Lysates were prepared in denaturing 2× LDS sample buffer (Invitrogen) with 1:10 β -mercaptoethanol to a final protein concentration of 2 $\mu\text{g } \mu\text{l}^{-1}$, and heated to 70 °C for 10 min before loading on gels. Ten micrograms of protein per lane was loaded for immunoblotting.

SCN organotypic slice culture experiments. Wild-type, ~10-day-old pups on a PER2-LUC (mice expressing a period2-luciferase fusion protein), C57BL/6 genetic background⁵⁶ were used for experiments. Brains were removed from pups and sectioned at 300 μm with a McIlwain Tissue Chopper. Slices were sorted and

trimmed to contain principally SCN tissue and placed onto a Millipore membrane insert (PICMORG) for culture at 37 °C in 5% CO₂ as described previously⁵⁷. Slices were transferred to 1.1 ml HEPES buffered medium with 100 μM beetle luciferin (Promega)⁵⁸ in a glass-bottomed Petri dish sealed with a coverslip and vacuum grease. Total bioluminescence was recorded with Hamamatsu photomultiplier tube assemblies housed in a light-tight 37 °C incubator, and recordings were expressed as counts per second integrated over 6-min sample bins⁵⁹. SCN slices were collected 'around the clock' every 4 h according to the phase of the PER2-LUC bioluminescence cycle. Circadian time (CT) 0 was operationally defined as the nadir in bioluminescence signal, and CT12 was taken to be at the peak. Individual SCN slices ($n = 3$ per time point) were immersed in 50 μl 2× LDS sample buffer (Invitrogen), and heated to 70 °C for 10 min before loading on gels. For immunoblotting, 10 μl of lysate per lane was loaded. Equal protein loading was confirmed by gel electrophoresis and Coomassie staining of gels loaded with equal volumes of lysate from each time point in each replicate set.

***Neurospora crassa*.** We used the following strains: wild type (*bd*, *matA* Fungal Genetics Stock Center (FGSC) accession 1858), *frq¹⁰* (*bd;frq¹⁰*, *matA* FGSC accession 7490), *frq⁷* (*bd;frq⁷*, *matA* FGSC accession 4898) and PRX-KO (NCU06031, *matA* FGSC accession 20012). All strains were maintained on Vogel's minimal media with 1.5% sucrose as a carbon source. Strain manipulation and growth media followed standard procedures⁶⁰. For circadian experiments, cultures were initiated by inoculating 10⁶ conidia in 25 ml of Vogel's medium containing 2% glucose, 0.5% arginine, 10 ng ml^{–1} biotin and 0.2% Tween 80. Plates were incubated under constant light for 36 h at 30 °C. Disks (1.2-cm diameter) were cut from the cohesive mycelial pad and three disks were placed in each of a series of 50 ml tubes (or 100 ml flasks) containing 30 ml (or 50 ml) of Vogel's medium with 0.03% glucose, 0.05% arginine and 10 ng ml^{–1} biotin⁶¹. These were incubated at 25 °C under constant light for at least 2 h before staggered transfers to constant darkness (25 °C) with shaking at 150 r.p.m. Mycelia were then collected at 4 h intervals, dried on paper, frozen in liquid N₂, and stored at –80 °C. Harvests were consolidated over 8 h to control for development/age of the tissue. All manipulations in the dark were performed under safe red light. For immunoblot analysis, tissue was ground in liquid N₂ with a mortar and pestle and suspended in ice-cold extraction buffer (50 mM HEPES, pH 7.4, 137 mM NaCl, 10% glycerol, 5 mM EDTA containing 50 $\mu\text{g ml}^{-1}$ phenylmethylsulphonylfluoride, 3 $\mu\text{g ml}^{-1}$ leupeptin and 3 $\mu\text{g ml}^{-1}$ pepstatin A⁶² and a phosphatase inhibitors cocktail (PhosSTOP, Roche) at a ratio of 0.5 ml of buffer to 0.5 ml of mycelia powder. For immunoblotting, an antiserum directed against human PRX6 (1-Cys) was used⁶³ (at 1:5,000 dilution) because no typical 2-Cys PRXs are annotated in the *N. crassa* genome so far, and thus antibody specificity could not otherwise be assured. Equal protein loading was confirmed by gel electrophoresis and Coomassie staining of gels loaded with equal volumes of lysate from every time point in each replicate set.

***Ostreococcus tauri*.** *O. tauri* cells were cultured as previously described⁹ and entrained in a 12 h light, 12 h dark cycle of blue light (17.5 $\mu\text{Em}^{-2}\text{s}^{-1}$, Lee lighting filter 724) at a constant temperature of 20 °C. Cultures of the arrhythmic TSL8 line (TOC1-overexpressing) and its parent line CCA1-LUC³⁰ were transferred into constant light at ZT0 and sampled every 4 h for 3 days. Five microlitres of cells was chilled on ice and pelleted at 4,500g at 4 °C for 10 min. The resulting pellet was resuspended in 50 μl sea water⁹, and cells were lysed by adding 50 μl 2× extraction buffer (Sigma-Aldrich, LUC-1 kit) and then 100 μl SDS loading buffer (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris-HCl, ~pH 6.8) while vortexing vigorously. Samples were heated to 100 °C for 5 min to denature proteins before loading on gels. Equal protein loading was confirmed by gel electrophoresis and Coomassie staining of gels loaded with equal volumes of lysate from every time point in each replicate set. For H₂O₂ treatment, cells were incubated for 30 min at the stated concentrations diluted in artificial sea water under normal culture conditions before lysis (described earlier).

***Synechococcus elongatus*.** The cyanobacterial strains used were *S. elongatus* sp. PCC7942 wild type (AMC149) and Δ KaiA mutant (AMC702). Cells were grown with aeration in constant light of 100 $\mu\text{E s}^{-1}\text{m}^{-2}$ at 30 °C to an attenuation (*D*) at 730 nm of 0.3. The $D_{730\text{ nm}}$ was maintained between 0.27 and 0.45 by dilution with fresh BG-11 medium^{64,65}. The culture was exposed to 12 h of constant darkness to synchronize the circadian clock, and then returned to constant light. At 4-h intervals under constant light, cells were collected, immediately frozen and then stored at –80 °C. Samples were prepared for immunoblotting as described previously⁶⁶. Equal protein loading was confirmed by gel electrophoresis and Coomassie staining of gels loaded with equal volumes of lysate from every time point in each replicate set. A single knockout for the 2-Cys *S. elongatus* sp. PCC7942 PRX gene⁶⁷ was generated by inserting an expression cassette for the kanamycin resistance gene into the EcoRI site near the amino-terminal coding region of the gene. For bioluminescence recordings, *psbA1p::luxAB* was used

as a reporter, using previously described protocols⁶⁸. The period of all the peroxiredoxin knockout strain showed no significant difference to that of wild-type. For H₂O₂ treatment, cells were incubated for 30 min at 1 mM H₂O₂ versus vehicle before cell collection.

Gel electrophoresis and western blotting. We used NuPAGE Novex 4–12% Bis-Tris gradient gels (Life Technologies), and ran them using the manufacturer's protocol with a non-reducing MES SDS buffer system, allowing characterization of proteins between 10 and 260 kDa^{9,10}. Protein transfer to nitrocellulose for blotting was performed using the iBlot system (Life Technologies), with a standard (programme P3, 7 min) protocol. Nitrocellulose was washed briefly, and then blocked for 30 min in 0.5% (w/w) BSA in non-fat dried milk (Marvel) in TBS containing 0.05% Tween-20 (TBST). After three brief washes in TBST, membranes were incubated in antibody diluted in blocking buffer (0.5% milk in BSA) overnight at 4 °C. The next day, membranes were washed in TBST for 5 min three times and then incubated with 1:10,000 horseradish peroxidase (HRP)-conjugated secondary antibody (Sigma-Aldrich) for 30 min. Four more 10-min washes were then performed before performing chemiluminescence detection using Immobilon reagent (Millipore), or ECL Plus reagent (GE Healthcare). To check even protein loading, the gels were stained with Coomassie SimplyBlue (Life Technologies). Antisera against peroxiredoxins were obtained from Abcam and used in blocking buffer (PRX-SO_{2/3} 1:10,000 dilution, and PRX1 1:2,000). Rabbit anti-Bmal1 antiserum was used at 1:2,000 in 0.5% BSA (Santa Cruz Antibodies). Rabbit anti-PER or anti-TIM was used at 1:10,000 dilution in blocking buffer. Anti-PER and anti-TIM were provided by R. Stanewsky and F. Rouyer, respectively. A mouse monoclonal β-actin antibody (Santa Cruz Antibodies) was used at 1:5,000 in 0.5% milk in BSA. Mouse anti-FRQ antibody was used at a dilution of 1:40 in blocking buffer. Anti-KaiC antiserum was used as described previously⁶⁶.

Phylogenetic analyses. We used the Mirrortree online server, using input FASTA sequences for human PRDX2 (GenBank accession CAG46588.1) and compared this serially to *S. elongatus* sp. PCC7942 proteins: KaiA (GenBank accession AAM82684.1), KaiB (GenBank accession AAM82685.1) and KaiC (AAM82686.1)⁶⁹. Similar analyses were performed for KaiC comparisons with other conserved bacterial proteins. Interspecies plots were generated, which contain a simplified representation of the correlation between the interprotein distances in phylogenetic trees for each protein being compared⁷⁰.

The Mirrortree web server implements the Mirrortree algorithm for calculating the tree similarity between two protein families, which has been demonstrated to be a good predictor of the interaction or functional relationships between them^{69,70}. Phylogenetic trees are obtained from these alignments with the neighbour-joining algorithm implemented in ClustalW⁷⁰ using bootstrap (100 repetitions) and excluding gaps for the calculation. The distance matrices are obtained by summing the branch lengths that separate each pair of proteins in the tree. Instead of calculating the complete matrices, only the proteins of organisms present in both trees are used. The similarity of trees between the two families is calculated as the correlation between their distance matrices according to the equation:

$$r = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}$$

in which n is the number of elements of the matrices, that is, $n = (N^2 - N)/2$; N is the number of common organisms; R_i is the element of the first distance matrix; S_i is the corresponding value for the second matrix; and \bar{R} and \bar{S} are the mean values of R_i and S_i , respectively. Correlation coefficients obtained for analyses of KaiC distance versus other conserved bacterial proteins (including KaiB, KaiC and PRX2) were compared by converting them to a normally distributed metric using Fisher's r -to- z transformation:

$$r' = \frac{1}{2} \ln \left| \frac{1+r}{1-r} \right|$$

in which r is the Mirrortree correlation coefficient; and r' is the Fisher-transformed correlation coefficient. Transformed coefficients (r') were then compared with each other to generate the test statistic, z :

$$z = \frac{r'_1 - r'_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

in which r'_1 is the first Fisher-transformed correlation coefficient; r'_2 is the second Fisher-transformed correlation coefficient; N_1 is number of common organisms in the first correlation; and N_2 is number of common organisms in the second correlation. We computed z with r'_1 representing the KaiC-PRX2 correlation coefficient, obtaining negative values for higher correlations, and positive values for lower ones (see Supplementary Table 8 and Supplementary Fig. 6). P values were then computed using standard tables (<http://faculty.vassar.edu/lowry/rdiff.html>).

Image and statistical analysis. Coomassie-stained gel images were obtained using a Licor Odyssey system, and immunoblot films were scanned using a back-illuminated flat-bed scanner. Densitometric quantification of images was performed using NIH ImageJ software. Signal was normalized against the respective loading control for each replicate at each time point for grouped data. Sine curve fitting was performed using Circwave software⁷¹, using a harmonic regression method, and ANOVA was also performed as an independent measure of temporal variation. Statistical comparisons between Mirrortree correlation coefficients were performed as detailed earlier.

- Gould, P. D. *et al.* Delayed fluorescence as a universal tool for the measurement of circadian rhythms in higher plants. *Plant J.* **58**, 893–901 (2009).
- Rosato, E. & Kyriacou, C. P. Analysis of locomotor activity rhythms in *Drosophila*. *Nature Protocols* **1**, 559–568 (2006).
- Reddy, A. B. *et al.* Circadian orchestration of the hepatic proteome. *Curr. Biol.* **16**, 1107–1115 (2006).
- Whitehead, K., Pan, M., Masumura, K., Bonneau, R. & Baliga, N. S. Diurnally entrained anticipatory behavior in archaea. *PLoS ONE* **4**, e5485 (2009).
- Mori, T., Binder, B. & Johnson, C. H. Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proc. Natl Acad. Sci. USA* **93**, 10183–10188 (1996).
- Yoo, S. H. *et al.* PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *Proc. Natl Acad. Sci. USA* **101**, 5339–5346 (2004).
- House, S. B., Thomas, A., Kusano, K. & Gainer, H. Stationary organotypic cultures of oxytocin and vasopressin magnocellular neurons from rat and mouse hypothalamus. *J. Neuroendocrinol.* **10**, 849–861 (1998).
- Hastings, M. H., Reddy, A. B., McMahon, D. G. & Maywood, E. S. Analysis of circadian mechanisms in the suprachiasmatic nucleus by transgenesis and biolistic transfection. *Methods Enzymol.* **393**, 579–592 (2005).
- Maywood, E. S. *et al.* Synchronization and maintenance of timekeeping in suprachiasmatic circadian clock cells by neuropeptidergic signaling. *Curr. Biol.* **16**, 599–605 (2006).
- Davis, R. H. *Neurospora: Contributions of a Model Organism* (Oxford Univ. Press, 2000).
- Morrow, M., Brunner, M. & Roenneberg, T. Assignment of circadian function for the *Neurospora* clock gene frequency. *Nature* **399**, 584–586 (1999).
- Olmedo, M. *et al.* A role in the regulation of transcription by light for RCO-1 and RCM-1, the *Neurospora* homologs of the yeast Tup1–Ssn6 repressor. *Fungal Genet. Biol.* **47**, 939–952 (2010).
- Woo, H. A. & Rhee, S. G. in *Methods in Redox Signaling* (ed. Das, D.) Ch. 4 19–23 (Mary Ann Liebert, 2010).
- Ishiyama, M. *et al.* Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. *Science* **281**, 1519–1523 (1998).
- Qin, X. *et al.* Intermolecular associations determine the dynamics of the circadian KaiABC oscillator. *Proc. Natl Acad. Sci. USA* **107**, 14805–14810 (2010).
- Xu, Y. *et al.* Intramolecular regulation of phosphorylation status of the circadian clock protein KaiC. *PLoS ONE* **4**, e7509 (2009).
- Stork, T., Laxa, M., Dietz, M. S. & Dietz, K. J. Functional characterisation of the peroxiredoxin gene family members of *Synechococcus elongatus* PCC 7942. *Arch. Microbiol.* **191**, 141–151 (2009).
- Xu, Y., Mori, T. & Johnson, C. H. Cyanobacterial circadian clockwork: roles of KaiA, KaiB and the *kaiBC* promoter in regulating KaiC. *EMBO J.* **22**, 2117–2126 (2003).
- Ochoa, D. & Pazos, F. Studying the co-evolution of protein families with the Mirrortree web server. *Bioinformatics* **26**, 1370–1371 (2010).
- Pazos, F. & Valencia, A. Similarity of phylogenetic trees as indicator of protein–protein interaction. *Protein Eng.* **14**, 609–614 (2001).
- Oster, H., Damerow, S., Hut, R. A. & Eichele, G. Transcriptional profiling in the adrenal gland reveals circadian regulation of hormone biosynthesis genes and nucleosome assembly genes. *J. Biol. Rhythms* **21**, 350–361 (2006).

Rab5 is necessary for the biogenesis of the endolysosomal system *in vivo*

Anja Zeigerer¹, Jerome Gilleron¹, Roman L. Bogorad², Giovanni Marsico¹, Hidenori Nonaka¹, Sarah Seifert¹, Hila Epstein-Barash³, Satya Kuchimanchi³, Chang Geng Peng³, Vera M. Ruda⁴, Perla Del Conte-Zerial¹, Jan G. Hengstler⁵, Yannis Kalaidzidis^{1,6}, Victor Koteliatsky^{2,3} & Marino Zerial¹

An outstanding question is how cells control the number and size of membrane organelles. The small GTPase Rab5 has been proposed to be a master regulator of endosome biogenesis. Here, to test this hypothesis, we developed a mathematical model of endosome dependency on Rab5 and validated it by titrating down all three Rab5 isoforms in adult mouse liver using state-of-the-art RNA interference technology. Unexpectedly, the endocytic system was resilient to depletion of Rab5 and collapsed only when Rab5 decreased to a critical level. Loss of Rab5 below this threshold caused a marked reduction in the number of early endosomes, late endosomes and lysosomes, associated with a block of low-density lipoprotein endocytosis. Loss of endosomes caused failure to deliver apical proteins to the bile canaliculi, suggesting a requirement for polarized cargo sorting. Our results demonstrate for the first time, to our knowledge, the role of Rab5 as an endosome organizer *in vivo* and reveal the resilience mechanisms of the endocytic system.

What are the molecular mechanisms underlying the biogenesis, number and size of cellular organelles? Organelles of the biosynthetic and endocytic pathway possess a toolbox of coat components, Rab GTPases, SNARE proteins, phosphoinositides and motors that collectively confer functional specificity^{1–5}. Whether any of these is a critical component for organelle biogenesis remains an unsolved problem. In the endocytic system, internalized cargo is transported through a network of hundreds of dynamically interacting early endosomes (EEs). These are pleomorphic organelles consisting of a vacuolar part, where cargo to be degraded accumulates, and tubular structures that deliver recycling cargo to the surface^{1,6}. Cargo destined for degradation is transferred from EEs to late endosomes via Rab conversion^{7,8} or vesicular transport⁹. The number of EEs depends on: (1) *de novo* generation (assuming that clathrin-coated vesicles (CCVs) can fuse homotypically rather than with pre-existing endosomes⁶), (2) homotypic fusion^{7,10,11} (reducing the number), (3) homotypic fission that generates new endosomes^{7,12}, and (4) early-to-late endosome conversion⁷ (decreasing the number). Homotypic fusion and fission events are common features of membrane organelles and the main determinants of EE number and size^{5,6}. Heterotypic fission that splits tubular structures containing recycling cargo from the vacuolar part can change the size but not the number of EEs. Endosomal trafficking also mediates the delivery of newly synthesized proteins to the plasma membrane^{13–15}, a function especially important for polarized cargo sorting in epithelial cells^{16,17}. Some apical and basolateral proteins are thought to pass through a recycling endosomal intermediate^{18–21}, but whether EEs are required for biosynthetic transport is unclear. As a central hub of a complex protein network governing membrane tethering, fusion and motility, the small GTPase Rab5 has been proposed to be a master regulator of EEs biogenesis²². This concept has received support from studies *in vitro*, where assembly of the Rab5 machinery on proteoliposomes with SNAREs led to efficient membrane fusion²³. Until now the role of Rab5 in endosome biogenesis *in vivo* has not been addressed as studies in mammalian cells and

Drosophila have concentrated on endocytosis and synaptic vesicles trafficking^{24–26}. Here, to address this outstanding problem, we focused on four main questions. Is Rab5 required for endosome biogenesis? If so, which consequences can decreasing Rab5 have for EEs? The levels of Rab5 may decline proportionally on all endosomes leading to endosome fragmentation^{10,27}. Alternatively, the levels of Rab5 per endosome may not change, whereas the total number of endosomes may decrease. Are EEs rate-limiting for late endocytic compartments? To what extent are EEs required for biosynthetic transport in polarized cells? To answer these questions we used a combination of mathematical modelling and *in vivo* short interfering RNA-based technology²⁸ to efficiently knockdown Rab5 in adult mouse liver.

Results

Modelling of endosome dependency on Rab5

To predict the dependency of the endocytic system on the levels of Rab5, we developed a simple mathematical model. We have previously described a model based on ordinary differential equations describing the Rab5 GDP/GTP cycle, effectors recruitment and Rab5-to-Rab7 conversion²⁹. Here, we expanded this model to include the number of endosomes (Methods and Supplementary Information, Box 1). The two components of the Rab5 cycle, shuttling between membrane and cytosol, and GDP/GTP exchange, were represented in a coarse-grained model. Single kinetic steps were grouped as non-linear rate dependencies on active Rab5, which could be monitored through the recruitment of its representative effector, EEA1. To model the number of endosomes, we added the Rab5-dependency of fusion and fission^{7,10–12,23}. Note that in this model the ratios of the kinetic rates define the system steady-state behaviour instead of individual parameters. In the model, we considered four main scenarios (Fig. 1, Supplementary Information and Supplementary Fig. 1):

(1) Both homotypic fusion and fission rates are independent of Rab5, implying that other components, for example, SNAREs³⁰, are rate-limiting. In this (negative control) case, the number of

¹Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. ²Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Amylin Pharmaceuticals, Inc., Cambridge, Massachusetts 02142, USA. ⁴Cardiovascular Research Center and Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. ⁵Leibniz Research Centre for Working Environment and Human Factors (IfADo), 44139 Dortmund, Germany. ⁶Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119899, Moscow, Russia.

BOX 1

Model coupling the Rab5 cycle to the endosome number

$$\begin{aligned}\frac{dRab5_{ee}}{dt} &= -Rab5_{ee} \cdot \frac{dN}{dt} \cdot \frac{1}{N} - k_{GAP}(Rab5_{ee}) \cdot Rab5_{ee} + k_{GEF}(Rab5_{ee}) \cdot rab5_{cyt} \\ \frac{drab5_{ee}}{dt} &= -rab5_{ee} \cdot \frac{dN}{dt} \cdot \frac{1}{N} - k_{GEF}(Rab5_{ee}) \cdot rab5_{ee} + k_{GAP}(Rab5_{ee}) \cdot Rab5_{ee} \\ &\quad + k_1 \cdot rab5_{cyt} - k_{-1} \cdot rab5_{ee} \\ \frac{drab5_{cyt}}{dt} &= -k_1 \cdot N \cdot rab5_{cyt} + k_{-1} \cdot rab5_{ee} \cdot N \\ \frac{dN}{dt} &= -k_{fus}(Rab5_{ee}) \cdot N^2 + k_{fis}(Rab5_{ee}) \cdot N\end{aligned}$$

Where: $Rab5_{ee}$, number of active Rab5 per endosome; $rab5_{ee}$, number of inactive Rab5 per endosome; $rab5_{cyt}$, total amount of inactive Rab5 in the cytosol; N , number of endosomes; $k_{GEF}(Rab5_{ee})$ and $k_{GAP}(Rab5_{ee})$, rates of Rab5 activation and inactivation; $k_{fus}(Rab5_{ee})$ and $k_{fis}(Rab5_{ee})$, endosome fusion and fission rates; k_1 , recruitment of Rab5 from cytosol; k_{-1} , extraction of Rab5 from early endosomes.

endosomes should not change (Fig. 1a, green line), whereas the amount of EEA1 per endosome would be reduced (Fig. 1b, green).

(2) The fusion but not fission rate depends on Rab5 or (3) the fusion rate depends on Rab5 and the fission rate is a power-law function of the fusion rate. Both lead to endosome fragmentation (Fig. 1a, blue and purple lines), whereas the amount of EEA1/endosome decrease even faster than in scenario 1 (Fig. 1b, blue and purple lines).

(4) Both fusion and fission rates are sigmoidal functions of Rab5. This scenario yields the most unexpected results. First, the number of endosomes demonstrates a non-monotonic dependency on Rab5 (Fig. 1a, red line), that is, the number drops in case of both increase (consistent with ref. 27) and decrease in Rab5. Second, the amount of EEA1/endosome shows a weak dependency on Rab5 over one order of magnitude (Fig. 1b, red line).

We next experimentally tested the predictions from the model. For this, a progressive reduction in Rab5 levels is preferable over gene knockout or loss-of-function mutations^{25–27}.

Hepatocyte-specific Rab5 knockdown in mouse liver

Most of the mechanistic understanding of Rab5 originated from cell-free systems^{10,11,23} and cultured cells^{7,27}. Rab5 has three isoforms, Rab5a,b,c²⁴, whose efficient depletion by RNAi in cultured cells causes cell cycle defects³¹, making results difficult to interpret. Therefore, we tested the role of Rab5 in endosome biogenesis *in vivo* in an organ where cells are mainly post-mitotic and within their physiological

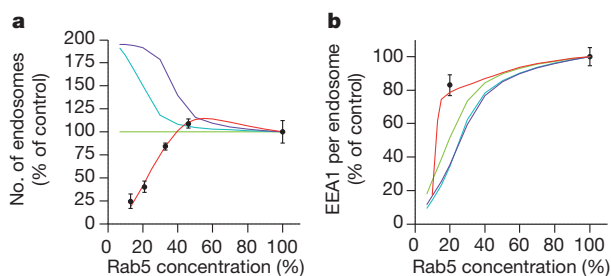


Figure 1 | Mathematical model of endosome number as function of Rab5 levels. a, b, Model predictions of number of endosomes (a) and amount of Rab5 effectors (for example, EEA1) per endosome as function of Rab5 levels for the four scenarios (b) (see text). Green line, scenario 1; blue line, scenario 2; purple line, scenario 3; red line, scenario 4. Black dots, experimental results (mean \pm s.e.m.).

environment. To this end, we took advantage of a novel technology based on lipid nanoparticles (LNPs) for specific delivery of siRNAs to mouse hepatocytes *in vivo*^{32,33}. Allowing the packaging of various siRNAs to achieve simultaneous silencing of multiple genes^{32,33}, this technology is ideal for knocking down all three Rab5 isoforms. Furthermore, the liver provides the opportunity to investigate the role of EEs in cell polarity *in vivo*. Specific Rab5a,b,c siRNAs were selected (Methods and Supplementary Fig. 2a), validated by 5'-RACE *in vivo* (Supplementary Fig. 2b) and incorporated as an equiponderant mixture (Rab5all) into LNPs optimized for liver-specific delivery^{32,33}. Control formulations consisted of the same concentration of Luciferase-targeting siRNAs. LNPs were delivered systemically via tail vein injection and displayed an *in vivo* dose response with maximal Rab5 messenger RNA silencing of approximately 80–90% for Rab5a, b, c at 0.5 mg kg⁻¹ each (Supplementary Fig. 2c). Rab5 knockdown (Rab5KD) was isoform- (Supplementary Fig. 2d) and liver-specific (Supplementary Fig. 2e). Maximal mRNA silencing of all Rab5 isoforms was achieved after 3 days and lasted at least 5 days, followed by a slow recovery (Fig. 2a). Importantly, LNPs did not produce any detectable inflammatory response or liver toxicity (Supplementary Fig. 3).

The reduction of Rab5 protein was determined by western blot analysis at 3–15 days post-injection using an anti-Rab5 pan antibody (Fig. 2b) recognizing all Rab5 isoforms equally (Supplementary Fig. 2f). We estimated the strongest reduction (~85%) 5 days post-injection (Fig. 2c). Because gene downregulation is mainly restricted to hepatocytes²⁸, the remaining levels of Rab5 could be accounted by non-parenchymal cells. Interestingly, at 10 days the levels of Rab5 protein partially recovered to approximately 35% of control.

Loss of EEA1 vesicles upon Rab5KD in liver

To determine the impact of Rab5 depletion on intracellular compartments in liver *in vivo*, it was necessary to optimize and further develop *ad hoc* staining protocols for each organelle marker in liver sections.

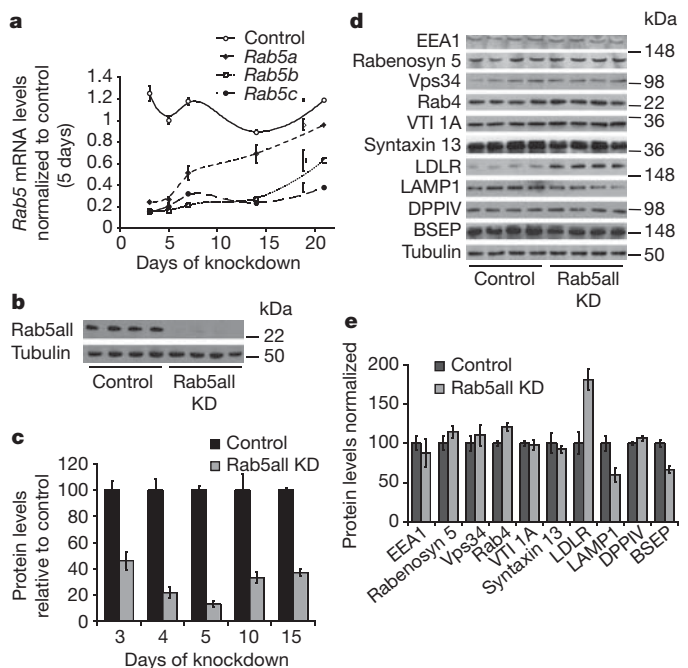


Figure 2 | Time course of Rab5 mRNA depletion and analysis of protein expression in mouse liver *in vivo*. a, Analysis of Rab5 mRNA levels over 21 days upon Rab5KD in mouse liver *in vivo* (mean \pm s.e.m., $n = 3$). b, Representative example of western blot analysis of Rab5 protein 5 days post injection ($n = 4$). c, Time course of Rab5 protein levels at the indicated times post injection by quantitative western blotting (mean \pm s.e.m.). d, Western blot analysis of endosomal machinery and cargo from liver lysates 5 days post-injection ($n = 4$). e, Quantification of western blots shown in d (mean \pm s.e.m.).

We first examined the staining pattern of the Rab5 effector and marker of EEs, EEA1 (ref. 34), 5 days after tail vein injection of LNPs. Using quantitative multi-parametric image analysis (QMPIA)³⁵, we measured a 70% reduction in number of EEA1-positive structures and total intensity of EEA1 on endosomes after Rab5 depletion compared to control (Fig. 3a, b). However, the mean intensity of EEA1/endosome was reduced by only 15%. The loss of EEA1-positive endosomes was not due to a reduction in EEA1 expression, as determined by protein analysis (Fig. 2d, e). Interestingly, knockdown of individual Rab5 isoforms had no effect on EEA1-positive endosomes (Supplementary Fig. 4), indicating that they are (partially) redundant, without excluding specific effects on cargo transport³⁶. These results are only consistent with scenario 4 predicting loss of EEs (Fig. 1a) but just a moderate decrease in EEA1 content per endosome (Fig. 1b).

Rab5KD causes loss of EEs, late endosomes and lysosomes

To verify that EE number is indeed reduced, we analysed liver sections by electron microscopy to compare the number of endosomes with the levels of Rab5 protein at each time point. In control sections, EE (large black arrow), late endosome/multi-vesicular bodies (MVBs; thin black arrow) and lysosomes (white arrow) could be discerned on the basis of their characteristic morphology^{37,38} (Fig. 4a). Three days post-injection, when the levels of Rab5 were already reduced by approximately 50%, the number of EEs was not significantly altered compared with control. In contrast, at days 4 and 5 when Rab5 was reduced to approximately 15% of total, the results were striking. First, there was a marked reduction in EEs (Fig. 4b), consistent with the decrease in EEA1-positive structures observed by light microscopy (Fig. 3). Interestingly, the mean size of remaining EEs after 4 ($246.3 \text{ nm} \pm 9.6$, P value = 0.2316) and 5 days ($224.8 \text{ nm} \pm 21.5$, P value = 0.1247) was only mildly reduced compared to control ($289.2 \text{ nm} \pm 37.9$). Second, we observed an approximately 80% reduction also in late endosomes/MVBs and lysosomes (Fig. 4c), reflecting the loss of the entire endolysosomal system. Interestingly, at 5 days there was a significant increase in clathrin-coated pits (CCPs) and clathrin-coated vesicles (CCVs) (Fig. 4d), and the basolateral plasma membrane was highly convoluted (Fig. 4b, middle inset), suggesting that endocytosis was inhibited (see below).

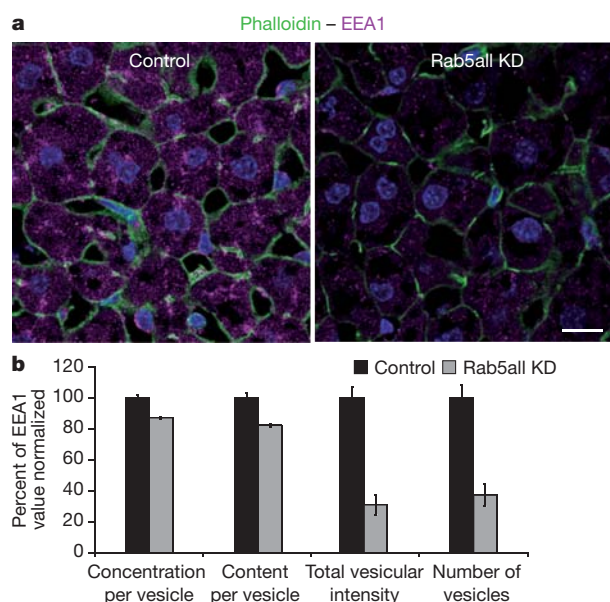


Figure 3 | Loss of EEA1 structures upon Rab5KD *in vivo*. **a**, Representative confocal microscopy images of liver sections stained with anti-EEA1 antibody, phalloidin and DAPI from control and Rab5KD animals (three merged middle confocal sections as maximal projections, see Methods in Supplementary Information). Scale bar, 20 μm . **b**, QMPIA of EEA1 vesicles from four different animals per condition (mean \pm s.e.m.).

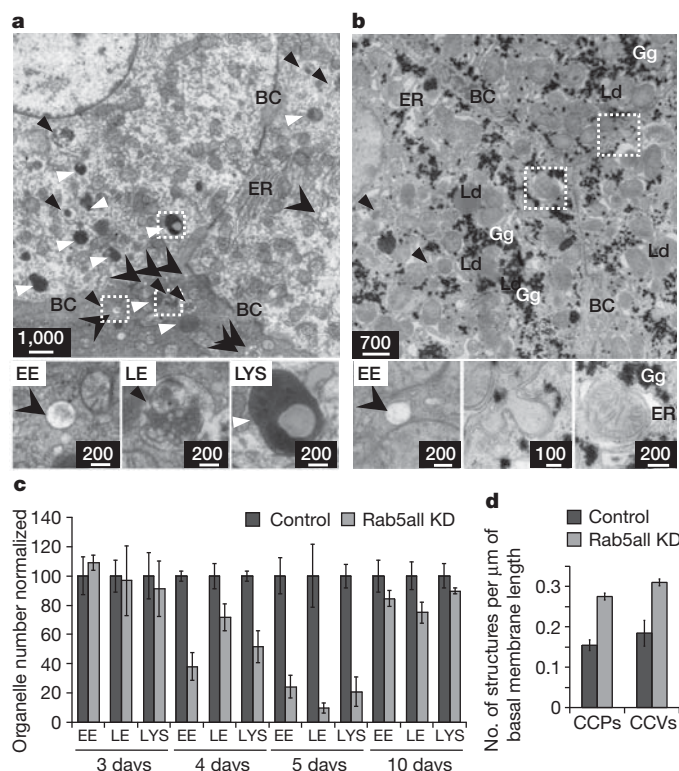


Figure 4 | Loss of early endosomes, MVBs and lysosomes upon Rab5KD.

a, b, Electron microscopy images of control (**a**) and KD (**b**) liver samples 5 days post-injection. BC, bile canaliculi; EE, early endosomes (large black arrow); LE, late endosomes (thin black arrow); LYS, lysosomes (white arrow); ER, endoplasmic reticulum; Gg, glycogen granules; Ld, lipid droplets. Insets show high magnification of representative structures. Convolution of the plasma membrane in Rab5KD (middle inset). Scale bars in nm. **c**, Time-course quantitative analysis of EE, LE and LYS profiles (mean \pm s.e.m.). **d**, Quantification of CCPs and CCVs normalized per μm of basolateral membrane 5 days post-injection (mean \pm s.e.m.).

Importantly, the phenotype was reversible. After 10 days, when Rab5 recovered to approximately 35% of control (Fig. 2c), the entire endocytic pathway was almost completely rebuilt (Fig. 4c). Notably, these data are entirely consistent with scenario 4 of the mathematical model, predicting changes in endosome number but not in amount of endosomal machinery (EEA1) per endosome (Fig. 3), in response to loss of Rab5 (Fig. 1a, b, black dots). In contrast to endosomes, the organelles of the biosynthetic pathway were much less affected upon Rab5KD. The relative serum levels of albumin and Factor-VII were normal (Supplementary Fig. 3d, e), indicative of functional secretion. We observed swelling of the endoplasmic reticulum (Fig. 4b, right inset), possibly reflecting metabolic perturbations³⁹. Indeed, hepatocytes contained numerous lipid droplets and glycogen granules, implying a severe metabolic dysfunction (Fig. 4b, right inset). No alterations of the Golgi complex were detectable by Giantin staining (Supplementary Fig. 5a, b). The number and total intensity of Rab11-positive structures, and the amount of Rab11/structure were only moderately reduced (15%) (Supplementary Fig. 5c, d) demonstrating that recycling endosomes do not disassemble despite loss of EEs. To explore the cause of depletion of endosome trafficking (Fig. 2d, e). We did not observe changes in EEA1, Rabenosyn-5, Vps34 and Rab4 in lysates of Rab5KD liver. The levels of the endosomal SNAREs Syntaxin-13 and VTI-1A were unaltered (Fig. 2d, e), indicating that the loss of endosomes is not consequent to the loss of expression of membrane tethering and fusion proteins²³. However, we detected a 50% decrease in LAMP1, consistent with a loss of late endocytic compartments (Fig. 2d, e and see below). Taken together, these data support the hypothesis that loss of endosomal

compartments is due to failure to assemble the transport machinery on membranes in the absence of Rab5.

Block of low-density lipoprotein uptake upon Rab5KD *in vitro*

The strong perturbations in EEs, late endosomes and lysosomes and the accumulation of CCPs and CCVs detected by EM upon Rab5 depletion prompted us to inspect the effects on endocytosis (Fig. 4d). Interestingly, we observed a strong increase in low-density lipoprotein (LDL) ($0.7015 \text{ mg ml}^{-1} \pm 0.211$, P value < 0.0001) in the serum of Rab5KD compared to control ($0.0700 \text{ mg ml}^{-1} \pm 0.039$) animals, arguing for reduced LDL uptake by the liver. To test this hypothesis directly, we measured LDL internalization kinetics. Given the difficulty of performing time-course uptake experiments *in vivo*, we used isolated primary mouse hepatocytes. Five days post-transfection with anti-Rab5all, we observed an approximately 80% reduction in mRNA levels of all Rab5 isoforms (Supplementary Fig. 2g). The number of EEA1-positive endosomes were reduced by about 70%, whereas the amount of EEA1/endosome was decreased only by about 15% (Supplementary Fig. 6a, c), corroborating the *in vivo* results (Fig. 3). LAMP1-positive structures (Supplementary Fig. 6b) and total vesicular intensity were also strongly reduced ($\sim 60\text{--}70\%$), again indicating loss of late endosomes/lysosomes (Supplementary Fig. 6d).

Next, we measured endocytosis of fluorescent LDL in primary mouse hepatocytes (Fig. 5). Five days post-transfection QMPIA revealed a marked slow-down in the kinetics of LDL uptake accompanied by a strong (70%) decrease in LDL-positive structures (Fig. 5a–c). These data therefore explain the high levels of LDL in the serum upon Rab5KD. Importantly, there was no downregulation of LDLR on the cell surface (Supplementary Fig. 6e). The total levels of LDLR were even increased *in vivo* (Fig. 2d, e), suggesting a possible compensatory mechanism to overcome the reduction in LDL uptake. To demonstrate that the phenotypic alterations in endosome number are specific for Rab5 depletion, we downregulated (90%) dynamin 2, a multidomain GTPase necessary for vesicle scission from the plasma membrane^{40,41} in primary mouse hepatocytes (Supplementary Fig. 7a,b). We observed an almost complete reduction in total LDL uptake and number of LDL vesicles (Fig. 5d, e) but, interestingly, no significant changes in number and intensity of EEA1- and LAMP1-positive structures (Fig. 5f).

Alterations in hepatocyte polarity *in vivo*

The ablation of EEs upon Rab5KD provided the opportunity to examine the dependence of polarized sorting on this endosomal compartment. We first inspected the polarity of hepatocytes by staining liver sections for the tight junction marker ZO1 (Supplementary Fig. 8, middle panels) and actin/phalloidin to detect cell borders and the bile canaliculi network (Supplementary Fig. 8, left panels). The latter was unchanged (Supplementary Fig. 8, right panels), ruling out disruption of cell polarity. We next inspected the localization of three apical proteins, dipeptidylpeptidase 4 (DPPIV), multidrug resistant-associated protein (MRP2) and the major bile salt export pump (BSEP), whose dependence on endosomal transport has never been demonstrated *in vivo*⁴². DPPIV was shown to follow the transcytotic route⁴³, whereas MRP2 trafficking remains controversial, being transcytosed⁴³ or following the direct apical targeting pathway⁴⁴. BSEP was suggested to pass through recycling endosomes⁴², although direct functional evidence for endosomal intermediates is lacking. In control liver, DPPIV was enriched in the bile canaliculi network where it colocalized with MRP2 (Fig. 6a, upper middle and right panel), with a small fraction on the basal side (Fig. 6a, upper left panel). Strikingly, in Rab5KD liver DPPIV redistributed to the basolateral surface of hepatocytes (Fig. 6a, lower left panel), colocalizing with CD166 and FcyR (Fig. 6b). DPPIV did not accumulate in Rab11-positive recycling endosomes or in the Golgi complex (Fig. 6b), and its expression was unaltered (Fig. 2d, e), indicating that after leaving the biosynthetic pathway DPPIV is trafficked to the basolateral side but fails to be re-internalized and sorted apically. In contrast, MRP2 correctly

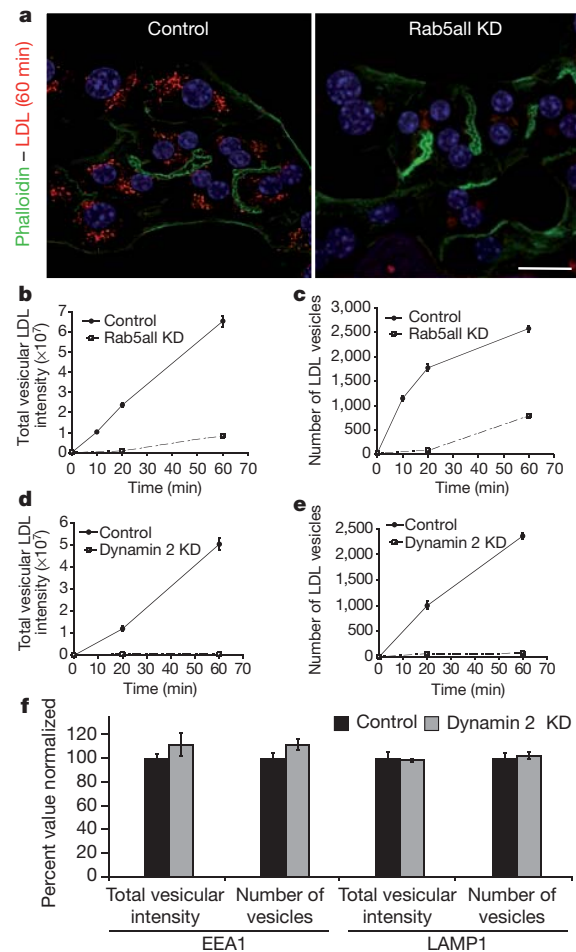


Figure 5 | Block of LDL endocytosis upon Rab5KD in primary hepatocytes *in vitro*. **a**, Representative confocal microscopy images of primary hepatocytes 5 days post-transfection, allowed to internalize DiI (1,1'-dioctadecyl-3,3,3'-tetramethylindocarbocyanine perchlorate)-LDL for various time points and stained with Alexa 488-phalloidin and DAPI. Scale bar, 20 μm . **b**, **c**, Time-course internalization of LDL analysed by QMPIA (key parameters are presented) (mean \pm s.e.m.). **d**–**f**, Downregulation of dynamin 2 in isolated primary mouse hepatocytes. **d**, **e**, LDL time-course data analysed by QMPIA (mean \pm s.e.m.). **f**, Confocal images of primary hepatocytes stained with anti-EEA1 and anti-LAMP1 antibodies and analysed by QMPIA (mean \pm s.e.m.).

localized to the apical surface (Fig. 6a, lower middle and right panel), suggesting that it follows a direct route. However, BSEP was greatly reduced ($\sim 70\%$) at the bile canaliculi (Fig. 6c, d) and redistributed intracellularly upon Rab5 depletion (Fig. 6c, inset right panel), indicating that it needs an endosomal intermediate for apical targeting. This strong decrease was not due to a similar reduction in protein expression, because BSEP protein levels were reduced only by approximately 30% (Fig. 2d, e). These results show that ablation of the endolysosomal system caused a strong impairment of transport of selected membrane proteins to the apical surface in hepatocytes *in vivo*. Altogether, our data highlight the functional role of the endocytic system in biosynthetic transport and cell polarity beyond what has been demonstrated so far.

Discussion

The most striking result of this study is that depletion of Rab5 *in vivo* led to loss of approximately 80% of the endolysosomal pathway, showing for the first time that Rab5 is a principal component of EE biogenesis *in vivo*. Other components, such as Rab5 effectors and SNAREs, continue to be expressed but, evidently, are not sufficient for providing structural and functional identity to EEs. Therefore,

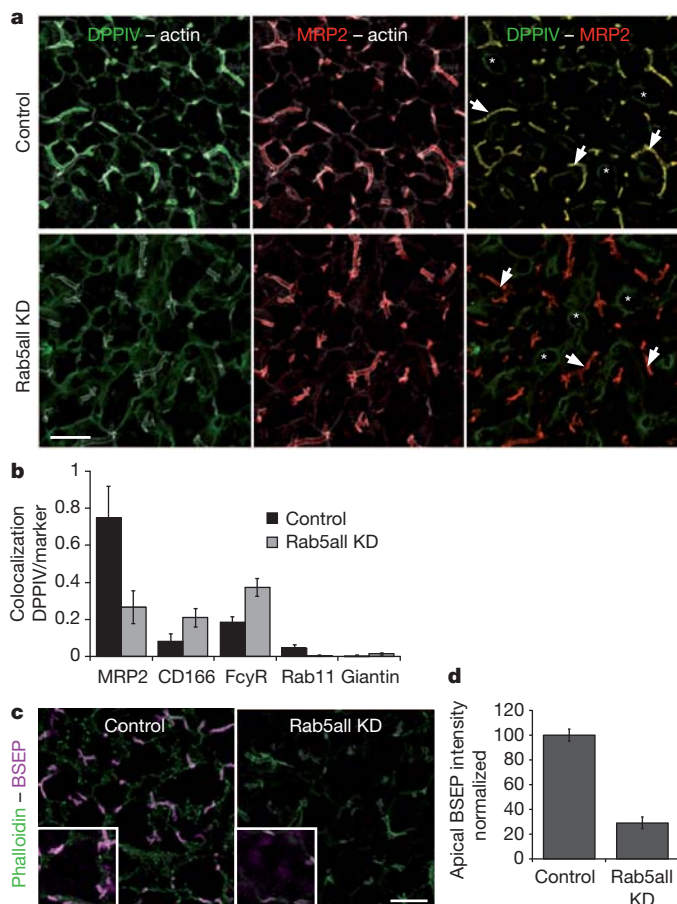


Figure 6 | Rab5KD causes mis-sorting of apical proteins in hepatocytes *in vivo*. **a**, Liver sections from control and Rab5KD animals were stained for DPPIV (left panels), MRP2 (middle panels) and actin. Merged images of DPPIV and MRP2 are shown in the right panels. Arrow, bile canaliculi; asterisk, sinusoids. Scale bar, 20 μ m. **b**, QMPIA of DPPIV distribution and colocalization with various markers ($n = 4$) (mean \pm s.e.m.). **c**, Images of liver sections stained for BSEP and actin. Scale bar, 20 μ m. Insets, high magnification of hepatocytes to visualize BSEP intracellular and apical localization in knockdown samples. **d**, Quantification of apical BSEP fluorescence intensity in control and Rab5KD liver (mean \pm s.e.m.).

Rab5 is mandatory for the assembly of the endosomal machinery and not a secondary regulator⁴⁵.

An important and unexpected finding is the resilience (defined, in the context of engineering and computing systems, as the ability to withstand and/or recover from strong perturbations) of the endocytic system to wide fluctuations in Rab5 concentrations. Loss of endosomes was not proportional to the levels of Rab5 but occurred abruptly upon around 80% reduction in protein levels (Fig. 2c), supporting only scenario 4 of the mathematical model (Fig. 1a, b). In this scenario, the fusion/fission dependency on Rab5 is a sigmoidal function, implying that wide variations in Rab5 levels result only in minor changes in Rab5 effectors per endosome. By analogy with engineering systems, Rab5 plays the role of the stabilizing element, that is, an element for which small changes are amplified and the feedback loops in the system (GDP/GTP exchange²⁹ and fusion/fission, Supplementary Fig. 1d, e) suppress variations. Consistent with the model, the finding that EEs were reduced in number, but not size, and retained a constant amount of effectors (Figs 1a, b and 3) suggests that each endosome requires a full complement of Rab5 machinery to be functional. The prediction of moderate decrease in endosome number upon increase in Rab5 is consistent with reported phenotypes²⁷. Interestingly, after 10 days Rab5 recovered to approximately 35% of control and the entire endocytic pathway was rebuilt. These data are

remarkably consistent with the model predicting even a moderate increase in endosome number with Rab5 reduced to 50% (Fig. 1a). For the interpretation of the phenotype it is important to consider EEs as a network in time rather than individual entities. Endosomes interact with one another within seconds⁷. As Rab5 gradually (over days) decreases upon RNAi, the endosomal network has time to adjust to a new steady state, with fewer endosomes being able to maintain normal amounts of Rab5 machinery. Consequently, EE biogenesis is suppressed and the EE network progressively shrinks. Because decrease in Rab5 reduces the rate of homotypic EE fusion^{7,10,11}, EEs would be expected to fragment. Yet, we found no evidence of fragmentation. Indeed, our model (Fig. 1) predicted that homotypic EE fission must be affected (possibly indirectly) even more than fusion below threshold (Supplementary Fig. 1b, c). This interpretation is supported by QMPIA of EE populations in LDL pulse-chase studies (Supplementary Information and Supplementary Fig. 9). As fewer EEs can reach the amount of Rab5 necessary for conversion^{7,29}, the number of late endocytic structures is reduced. This effect is therefore secondary to the loss of EEs. Note that the residual endosomes in the network are still functional, capable of receiving and propagating cargo for degradation (Supplementary Information and Supplementary Fig. 9b). Upon Rab5KD, there may be effects on the heterotypic fission of recycling tubules and vesicles as well. However, Rab11 compartments were reduced in number only by 15% (Supplementary Fig. 5c, d). This mild effect and the surface levels of LDLR being also unchanged (Supplementary Fig. 6e) suggest that recycling may not be dysfunctional (see Supplementary Information).

Given the time scale of the effect (5 days), the ‘disappearance’ of lysosomes is surprising, because indigestible substances accumulate within lysosomes for weeks⁴⁶. However, lysosomes within hepatocytes are able to release their content into the bile canaliculi by exocytosis⁴⁶ and, therefore, such organelles may not have been replaced.

An unresolved problem is whether vesicles derived from the plasma membrane contribute to the biogenesis of EEs⁶. Blockage of dynamin 2-dependent endocytosis did not alter the number of endosomes (Fig. 5f), which is inconsistent with an obligatory role of endocytic vesicles in EEs biogenesis. However, caution in the interpretation is suggested by the possible induction of dynamin-independent compensatory endocytic mechanisms (clathrin-independent carriers (CLICs)/GPI-enriched endocytic compartments (GEECs), Arf6-dependent) that may maintain EEs⁴⁷. Their onset may require time, unlike the use of temperature-sensitive mutants⁴⁸.

Liver function depends on the polarity of hepatocytes. Upon loss of Rab5 the delivery of DPPIV and BSEP to the bile canaliculi was strongly impaired (Fig. 6). In contrast, the localization of MRP2, which was postulated to follow a transcytotic route⁴³ to the bile canaliculi, was unaffected, supporting direct targeting for this transporter⁴⁴. Subcellular fractionation studies indicated that apical proteins traverse endosomal compartments, primarily Rab11-positive recycling endosomes, on their way to the bile canaliculi^{42,49}. Our data, however, suggest that EEs are also required as sorting and transport intermediates for the delivery of certain proteins to the apical surface in hepatocytes. Overall, the primary cells and organ system used in this study highlight the importance of using a physiological system to investigate cell polarity and tissue morphogenesis.

METHODS SUMMARY

The following methods are described in detail in Supplementary Information. Modelling was done by simulation and fitting ordinary differential equations (ODEs) using FitModel. LNP siRNA formulations^{32,33} were administered to C57BL/6N mice through tail vein injection. Livers were collected after various time points and analysed for mRNA (quantitative PCR with reverse transcription) and protein determination (western blotting), and sectioned for immunocytochemistry and electron microscopy. Western blots were quantified with ImageJ (<http://lukemiller.org/index.php/2010/11/analyzing-gels-and-western-blots-with-image-j/>). Immunocytochemistry was performed on optimal cutting temperature compound (OCT)-embedded liver cryosections. For electron

microscopy, liver tissues were fixed, embedded and sectioned following classical procedures. Primary mouse hepatocytes were isolated, cultured and subjected to immunofluorescence analysis. LDL internalization assays were performed as described⁷. QMPIA was performed as described³⁵.

Received 13 December 2011; accepted 3 April 2012.

- Mellman, I. Endocytosis and molecular sorting. *Annu. Rev. Cell Dev. Biol.* **12**, 575–625 (1996).
- Bonifacino, J. S. & Glick, B. S. The mechanisms of vesicle budding and fusion. *Cell* **116**, 153–166 (2004).
- Robinson, M. S. Adaptable adaptors for coated vesicles. *Trends Cell Biol.* **14**, 167–174 (2004).
- Jahn, R. & Scheller, R. H. SNAREs—engines for membrane fusion. *Nature Rev. Mol. Cell Biol.* **7**, 631–643 (2006).
- Pfeffer, S. R. Unsolved mysteries in membrane traffic. *Annu. Rev. Biochem.* **76**, 629–645 (2007).
- Gruenberg, J. & Maxfield, F. R. Membrane transport in the endocytic pathway. *Curr. Opin. Cell Biol.* **7**, 552–563 (1995).
- Rink, J., Ghigo, E., Kalaidzidis, Y. & Zerial, M. Rab conversion as a mechanism of progression from early to late endosomes. *Cell* **122**, 735–749 (2005).
- Poteryaev, D., Datta, S., Ackema, K., Zerial, M. & Spang, A. Identification of the switch in early-to-late endosome transition. *Cell* **141**, 497–508 (2010).
- Vonderheit, A. & Helenius, A. Rab7 associates with early endosomes to mediate sorting and transport of Semliki forest virus to late endosomes. *PLoS Biol.* **3**, e233 (2005).
- Gorvel, J. P., Chavrier, P., Zerial, M. & Gruenberg, J. rab5 controls early endosome fusion *in vitro*. *Cell* **64**, 915–925 (1991).
- Rybin, V. *et al.* GTPase activity of Rab5 acts as a timer for endocytic membrane fusion. *Nature* **383**, 266–269 (1996).
- Skjeldal, F. M. *et al.* The fusion of early endosomes induces molecular motor-driven tubule formation and fission. *J. Cell Science* <http://dx.doi.org/10.1242/jcs.092569> (22 February 2012).
- Futter, C. E., Connolly, C. N., Cutler, D. F. & Hopkins, C. R. Newly synthesized transferrin receptors can be detected in the endosome before they appear on the cell surface. *J. Biol. Chem.* **270**, 10999–11003 (1995).
- Harsay, E. & Schekman, R. A subset of yeast vacuolar protein sorting mutants is blocked in one branch of the exocytic pathway. *J. Cell Biol.* **156**, 271–286 (2002).
- Ang, A. L. *et al.* Recycling endosomes can serve as intermediates during transport from the Golgi to the plasma membrane of MDCK cells. *J. Cell Biol.* **167**, 531–543 (2004).
- Weisz, O. A. & Rodriguez-Boulant, E. Apical trafficking in epithelial cells: signals, clusters and motors. *J. Cell Sci.* **122**, 4253–4266 (2009).
- Golachowska, M. R., Hoekstra, D. & van IJendoorn, S. C. D. Recycling endosomes in apical plasma membrane domain formation and epithelial cell polarity. *Trends Cell Biol.* **20**, 618–626 (2010).
- Ihrke, G. *et al.* Apical plasma membrane proteins and endolyn-78 travel through a subapical compartment in polarized WIF-B hepatocytes. *J. Cell Biol.* **141**, 115–133 (1998).
- Cresawn, K. O. *et al.* Differential involvement of endocytic compartments in the biosynthetic traffic of apical proteins. *EMBO J.* **26**, 3737–3748 (2007).
- Nokes, R. L., Fields, I. C., Collins, R. N. & Folsch, H. Rab13 regulates membrane trafficking between TGN and recycling endosomes in polarized epithelial cells. *J. Cell Biol.* **182**, 845–853 (2008).
- Farr, G. A., Hull, M., Mellman, I. & Caplan, M. J. Membrane proteins follow multiple pathways to the basolateral cell surface in polarized epithelial cells. *J. Cell Biol.* **186**, 269–282 (2009).
- Zerial, M. & McBride, H. Rab proteins as membrane organizers. *Nature Rev. Mol. Cell Biol.* **2**, 107–117 (2001).
- Ohya, T. *et al.* Reconstitution of Rab- and SNARE-dependent membrane fusion by synthetic endosomes. *Nature* **459**, 1091–1097 (2009).
- Bucci, C. *et al.* Co-operative regulation of endocytosis by three Rab5 isoforms. *FEBS Lett.* **366**, 65–71 (1995).
- Wucherpfennig, T., Wilsch-Brauninger, M. & Gonzalez-Gaitan, M. Role of *Drosophila* Rab5 during endosomal trafficking at the synapse and evoked neurotransmitter release. *J. Cell Biol.* **161**, 609–624 (2003).
- Morrison, H. A. *et al.* Regulation of early endosomal entry by the *Drosophila* tumor suppressors Rabenosyn and Vps45. *Mol. Biol. Cell* **19**, 4167–4176 (2008).
- Bucci, C. *et al.* The small GTPase rab5 functions as a regulatory factor in the early endocytic pathway. *Cell* **70**, 715–728 (1992).
- Akinc, A. *et al.* Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. *Mol. Ther.* **18**, 1357–1364 (2010).
- Del Conte-Zerial, P. *et al.* Membrane identity and GTPase cascades regulated by toggle and cut-out switches. *Mol. Syst. Biol.* **4**, 206 (2008).
- Heinrich, R. & Rapoport, T. A. Generation of nonidentical compartments in vesicular transport systems. *J. Cell Biol.* **168**, 271–280 (2005).
- Serio, G. *et al.* Small GTPase Rab5 participates in chromosome congression and regulates localization of the centromere-associated protein CENP-F to kinetochores. *Proc. Natl Acad. Sci. USA* **108**, 17337–17342 (2011).
- Akinc, A. *et al.* A combinatorial library of lipid-like materials for delivery of RNAi therapeutics. *Nature Biotechnol.* **26**, 561–569 (2008).
- Love, K. T. *et al.* Lipid-like materials for low-dose, *in vivo* gene silencing. *Proc. Natl Acad. Sci. USA* **107**, 1864–1869 (2010).
- Christoforidis, S., McBride, H. M., Burgoyne, R. D. & Zerial, M. The Rab5 effector EEA1 is a core component of endosome docking. *Nature* **397**, 621–625 (1999).
- Collinet, C. *et al.* Systems survey of endocytosis by multiparametric image analysis. *Nature* **464**, 243–249 (2010).
- Chen, P. I., Kong, C., Su, X. & Stahl, P. D. Rab5 isoforms differentially regulate the trafficking and degradation of epidermal growth factor receptors. *J. Biol. Chem.* **284**, 30328–30338 (2009).
- Mukherjee, S., Ghosh, R. N. & Maxfield, F. R. Endocytosis. *Physiol. Rev.* **77**, 759–803 (1997).
- Schroeder, B. M. M. in *The Liver: Biology and Pathobiology* 5th edn (eds Arias, I. *et al.*) Ch. 7 107–123 (Wiley-Blackwell, 2009).
- Wolfsdorf, J. I. & Weinstein, D. A. Glycogen storage diseases. *Rev. Endocr. Metab. Disord.* **4**, 95–102 (2003).
- van der Bliek, A. M. *et al.* Mutations in human dynamin block an intermediate stage in coated vesicle formation. *J. Cell Biol.* **122**, 553–563 (1993).
- Shen, H. *et al.* Constitutive activated Cdc42-associated kinase (Ack) phosphorylates at arrested endocytic clathrin-coated pits of cells that lack dynamin. *Mol. Biol. Cell* **22**, 493–502 (2011).
- Kipp, H., Pichetshote, N. & Arias, I. M. Transporters on demand: intrahepatic pools of canalicular ATP binding cassette transporters in rat liver. *J. Biol. Chem.* **276**, 7218–7224 (2001).
- Wang, L. & Boyer, J. L. The maintenance and generation of membrane polarity in hepatocytes. *Hepatology* **39**, 892–899 (2004).
- Wakabayashi, Y., Dutt, P., Lippincott-Schwartz, J. & Arias, I. M. Rab11a and myosin Vb are required for bile canalicular formation in WIF-B9 cells. *Proc. Natl Acad. Sci. USA* **102**, 15087–15092 (2005).
- Lawe, D. C. *et al.* Sequential roles for phosphatidylinositol 3-phosphate and Rab5 in tethering and fusion of early endosomes via their interaction with EEA1. *J. Biol. Chem.* **277**, 8611–8617 (2002).
- Renaud, G., Hamilton, R. L. & Havel, R. J. Hepatic metabolism of colloidal gold-low-density lipoprotein complexes in the rat: evidence for bulk excretion of lysosomal contents into bile. *Hepatology* **9**, 380–392 (1989).
- Damke, H., Baba, T., van der Bliek, A. M. & Schmid, S. L. Clathrin-independent pinocytosis is induced in cells overexpressing a temperature-sensitive mutant of dynamin. *J. Cell Biol.* **131**, 69–80 (1995).
- Koenig, J. H. & Ikeda, K. Transformational process of the endosomal compartment in nephrocytes of *Drosophila melanogaster*. *Cell Tissue Res.* **262**, 233–244 (1990).
- Wakabayashi, Y., Lippincott-Schwartz, J. & Arias, I. M. Intracellular trafficking of bile salt export pump (ABCB11) in polarized hepatic cells: constitutive cycling between the canalicular membrane and rab11-positive endosomes. *Mol. Biol. Cell* **15**, 3485–3496 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge K. Simons, E. Knust, C. Fetzer, T. Galvez, G. O'Sullivan and M. P. McShane for discussions and comments on the manuscript. We thank W. John and A. Muench-Wuttke from the Biomedical Services Facility for mouse care and injections. We acknowledge J. Peychl for the management of the Light Microscopy Facility and K. Manygoats as well as J.-M. Verbavatz for their support. We thank B. Bettencourt and J. Hettinger for siRNA design and serum biochemistry assays, respectively. This work was financially supported by the Virtual Liver initiative (<http://www.virtual-liver.de>), funded by the German Federal Ministry of Research and Education (BMBF), the Max Planck Society (MPG) and the DFG. A.Z. was supported by a grant from Marie Curie Action, Intra-European Fellowship (fp7-people-ief-2008) and J.G. from an EMBO long-term fellowship.

Author Contributions M.Z. and V.K. conceived the project and M.Z. directed it. A.Z. designed and directed the animal injection experiments, and performed the stainings and imaging of liver sections. J.G. and A.Z. co-developed the staining procedures for the tissue sections. J.G.H. helped A.Z. to establish the hepatocyte isolation technique. A.Z. established the primary culture, developed the endocytosis assays, staining protocols and the knockdown technique with LNPs in primary hepatocytes. J.G. performed the electron microscopy analysis and quantifications, and the sectioning of liver tissue. R.L.B. selected siRNAs and validated their efficacy and specificity *in vitro* and *in vivo*, and designed, performed and analysed some *in vivo* experiments. H.E.-B. prepared the siRNAs into lipidoid-based formulations and analysed them. S.K. and C.G.P. performed the synthesis and analysis of siRNAs. V.M.R. designed, performed and analysed the 5'RACE assay. H.N. performed the RT-PCR in primary hepatocytes. S.S. under the supervision of A.Z. and J.G. performed the western blot analysis and performed the hepatocyte isolation and primary culturing under the supervision of A.Z.; G.M. under the supervision of Y.K. adapted the QMPIA for primary hepatocytes and liver tissue and performed the image analysis. Y.K. and P.d.C.-Z. developed the mathematical model. M.Z., Y.K., A.Z. and J.G. wrote the manuscript, R.L.B. and V.K. participated in the editing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.Z. (zerial@mpi-cbg.de).

Superflares on solar-type stars

Hiroyuki Maehara¹, Takuya Shibayama¹, Shota Notsu¹, Yuta Notsu¹, Takashi Nagao¹, Satoshi Kusaba¹, Satoshi Honda¹, Daisaku Nogami¹ & Kazunari Shibata¹

Solar flares are caused by the sudden release of magnetic energy stored near sunspots. They release 10^{29} to 10^{32} ergs of energy on a timescale of hours¹. Similar flares have been observed on many stars, with larger ‘superflares’ seen on a variety of stars^{2,3}, some of which are rapidly rotating^{4,5} and some of which are of ordinary solar type^{3,6}. The small number of superflares observed on solar-type stars has hitherto precluded a detailed study of them. Here we report observations of 365 superflares, including some from slowly rotating solar-type stars, from about 83,000 stars observed over 120 days. Quasi-periodic brightness modulations observed in the solar-type stars suggest that they have much larger starspots than does the Sun. The maximum energy of the flare is not correlated with the stellar rotation period, but the data suggest that superflares occur more frequently on rapidly rotating stars. It has been proposed that hot Jupiters may be important in the generation of superflares on solar-type stars⁷, but none have been discovered around the stars that we have studied, indicating that hot Jupiters associated with superflares are rare.

We searched for stellar flares on solar-type stars (G-type main-sequence stars) using data collected by NASA’s Kepler⁸ mission during the period from April 2009 to December 2009 (a brief summary of our flare search method is described in the legend of Fig. 1 and more detail is provided in Supplementary Information). We used the effective temperature (T_{eff}) and the surface gravity ($\log(g)$) available from the Kepler Input Catalog⁹ to select solar-type stars. The selection criteria are as follows: $5,100 \text{ K} \leq T_{\text{eff}} < 6,000 \text{ K}$, $\log(g) \geq 4.0$. The total numbers of solar-type stars are 9,751 for quarter 0 of the Kepler mission (the length of observation period is about 10 d), 75,728 for quarter 1 (33 d), 83,094 for quarter 2 (90 d) and 3,691 for quarter 3 (90 d).

We found 365 superflares (flares with energy $>10^{33}$ erg) on 148 solar-type stars (light curves of each flare are summarized in Supplementary Fig. 8 and properties of each flare star are listed in Supplementary Table 1). The durations of the detected superflares are typically a few hours, and their amplitudes are generally of order 0.1–1% of the stellar luminosity. The bolometric luminosity and total bolometric energy of each flare were estimated from the stellar radius, the effective temperature in the Kepler Input Catalog, and the observed amplitude and duration of the flare by assuming that the spectrum of white-light flares can be described by blackbody radiation with an effective temperature of 10,000 K (ref. 10). We considered the spectral response of the Kepler photometer in doing our luminosity and energy calculations. The bolometric luminosity of superflares on G-type main-sequence stars ranges from 9×10^{29} to $4 \times 10^{32} \text{ erg s}^{-1}$, and the total bolometric energy of superflares ranges from 10^{33} to 10^{36} erg (hereafter the luminosities and energies of flares given are the bolometric values). The uncertainties in the luminosities and energies of flares are estimated to be about $\pm 60\%$.

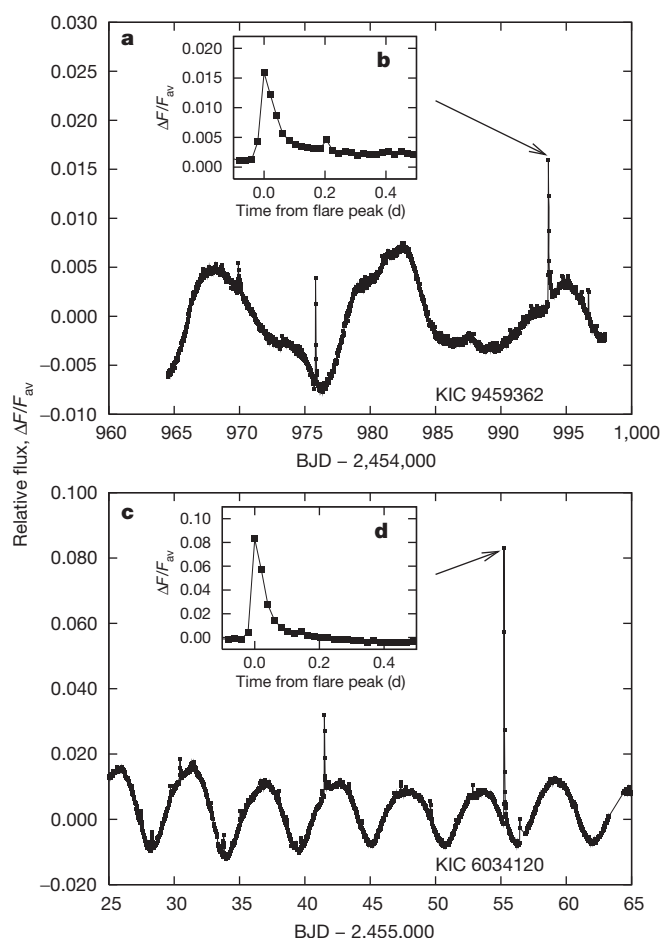


Figure 1 | Light curve of typical superflares. **a**, Light curve of superflares on the G-type main-sequence star KIC 9459362. The individual points represent the difference between the observed brightness during each cadence and the average brightness during the observation period. The typical photometric error is about 0.02%. BJD, barycentric Julian date. **b**, Enlarged light curve of a superflare observed at BJD 2,454,993.63. The relative flux ($\Delta F/F_{\text{av}}$) and the duration of this flare are respectively 1.4% and 3.9 h, which correspond to a peak bolometric luminosity of $1.3 \times 10^{31} \text{ erg s}^{-1}$ and a total energy of 5.6×10^{34} erg. **c**, Same as **a**, but for the G-type main-sequence star KIC 6034120. **d**, Same as **b**, but for the superflare that occurred on KIC 6034120 at BJD 2,455,055.22. The relative flux and the duration of this flare are respectively 8.4% and 5.4 h, which correspond to a peak bolometric luminosity of $6.8 \times 10^{31} \text{ erg s}^{-1}$ and a total energy of 3.0×10^{35} erg. These superflares were detected by the following procedure. First we analysed the long-cadence calibrated flux (time resolution, 29.4 min) and calculated the distribution of brightness changes between all pairs of consecutive data points. Then the threshold of the flare was determined to be three times the value in the top 1% of the distribution. The threshold value is typically about 0.1% of the brightness of the star. Finally we defined as flare candidates those data points at which the brightness change exceeded the threshold. Because many of these flare candidates are false events (for various reasons), for confirmation we examined their durations and the shapes of their light curves, checked for the existence of nearby stars and studied the charge-coupled-device pixel-level data for each flare candidate.

¹Kwasan and Hida Observatories, Kyoto University, 17 Ohmine-cho Kita Kazan, Yamashina-ku, Kyoto City, Kyoto 607-8471, Japan.

The brightnesses of the two stars in Fig. 1 show quasi-periodic variation. The period of the variation of KIC 9459362 is 12.5 d and that of KIC 6034120 is 5.7 d. The brightness variation can be caused by one of the following mechanisms: rotation of a star with starspots¹¹, orbital motion of a binary system¹², eclipse by an accompanying star¹² or stellar pulsation¹³. Here stellar pulsation can be excluded because the pulsation period of the G-type main-sequence stars is shorter than a few hours¹³. The possibility of the brightness variation being due to rotation must be carefully distinguished from its being due to orbital motion, on the basis of the difference in the shape of the light curves. The variability type of KIC 9459362 and KIC 6034120 is classified as stellar rotational modulation¹⁴. The energy of superflares can be explained by the magnetic energy stored near starspots. If the time variations in the brightnesses of KIC 9459362 and KIC 6034120 are caused by their rotation and the existence of starspots, their brightness variations suggest that solar-type stars with superflares have much larger starspots than does the Sun (for more details on the relation between starspots and superflares, see Supplementary Information). Typical periods of brightness variation range from one day to a few tens of days. According to the results of the variability-type classification¹⁴, about 65% of the brightness variations in the quarter-1 data can be taken to result from stellar rotational modulation. Although we could not exclude the possibility of the binary-eclipse scenario by using only photometric data, we assumed that the period of quasi-periodic modulation corresponds to the period of stellar rotation (for more

details on the possibility of superflares on secondary stars, see Supplementary Information).

The average occurrence frequency of superflares can be estimated from the number of observed superflares, the number of observed stars and the length of the observation period. For example, in the case of slowly rotating G-type main-sequence stars with surface temperatures of $5,600 \text{ K} \leq T_{\text{eff}} < 6,000 \text{ K}$, 14 superflares were detected from the data on about 14,000 stars over 120 d. Hence, the occurrence frequency of superflares is 2.9×10^{-3} flares per year per star, which corresponds to a superflare occurring on a star once every 350 yr. The occurrence frequency distribution of superflares on solar-type stars can be fitted in the energy range $\geq 4 \times 10^{34} \text{ erg}$ using a simple power law (Fig. 2a, b). The frequency distribution function of superflares on solar-type stars is similar to those of solar flares¹⁵ and stellar flares on red dwarfs¹⁶. The power-law index of the distribution of superflares (-2.0 to -2.3) is nearly equal to that of the distribution of solar flares. The occurrence frequency of superflares on slowly rotating G-type main-sequence stars is about ten times lower than the average occurrence frequency (Fig. 2b). The occurrence frequency also depends on the surface temperature of the star (Fig. 2c, d), and is higher for lower-temperature ($5,100 \text{ K} \leq T_{\text{eff}} < 5,600 \text{ K}$) G-type main-sequence stars than for higher-temperature ($5,600 \text{ K} \leq T_{\text{eff}} < 6,000 \text{ K}$) stars. The superflares of 10^{34} erg on Sun-like stars (that is, slowly rotating G-type main-sequence stars with surface temperatures of $5,600 \text{ K} \leq T_{\text{eff}} < 6,000 \text{ K}$; Fig. 2d) occur once every 800 yr, and those of 10^{35} erg occur once every

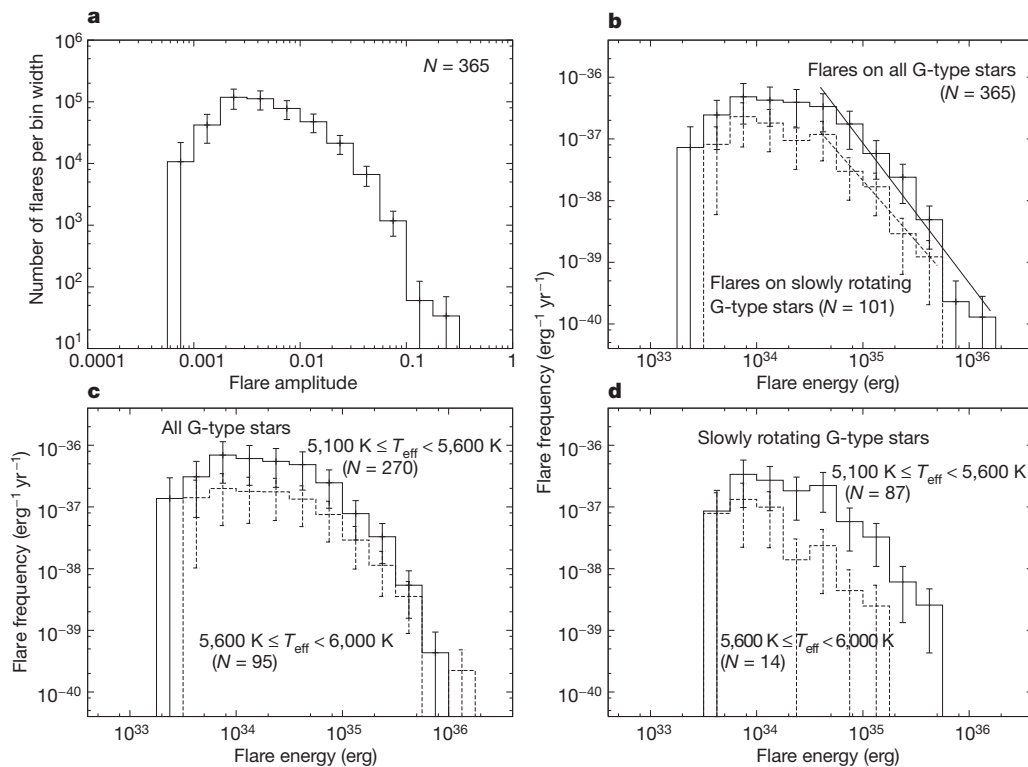


Figure 2 | Frequency distribution of superflares on G-type main-sequence stars. **a**, Number distribution of flares as a function of observed flare amplitude. The errors were estimated from the 10% uncertainty in the amplitude and the square root of the event number in each bin. The number of flares (N) is 365. **b**, Frequency distribution of flares as a function of flare energy. The vertical axis indicates the number of superflares per star per year per unit energy. The histograms show the frequency distribution of flares on all G-type main-sequence stars (solid line) and on G-type main-sequence stars excluding the stars that have quasi-periodic brightness modulations with periods shorter than 10 d (dashed line). The error bars represent the 1σ uncertainty in the frequency estimated from the uncertainty in the energy estimation and the square root of the event number. The occurrence distribution of superflares can be fitted by a

simple power-law function in the large-energy regime between 4×10^{34} and $2 \times 10^{36} \text{ erg}$. The power-law indexes of the distributions are -2.3 ± 0.3 and -2.0 ± 0.2 , respectively (downward-sloping lines). Because the detection completeness of flares with energy less than $4 \times 10^{34} \text{ erg}$ is low (10^{-3} for 10^{33} erg and 10^{-1} for 10^{34} erg), we did not use these data for the fitting. Because of the small number of events in the energy bins above 10^{36} erg , the fit does not well represent the data. **c**, Same as **b**, but for flares on all lower-temperature ($5,100 \text{ K} \leq T_{\text{eff}} < 5,600 \text{ K}$; solid line) stars and all higher-temperature ($5,600 \text{ K} \leq T_{\text{eff}} < 6,000 \text{ K}$; dashed line) stars. **d**, Same as **c**, but only for flares on slowly rotating stars (that is, excluding stars showing brightness variations with periods shorter than 10 d).

5,000 yr, although accurate statistics are difficult to obtain because only 14 superflares have been observed on Sun-like stars.

As shown in Fig. 3a, the maximum energy of a superflare does not show any clear correlation with the period of stellar rotation, assuming that the period of brightness modulation corresponds to the rotational period of the star. If the flare energy can be explained by the magnetic energy stored near the starspot¹¹, this result suggests that the maximum magnetic energy stored near the spot does not have a strong dependence on the period of rotation. This result also implies that superflares can occur on slowly rotating solar-type stars like the Sun.

The frequency of superflares tends to decrease as the period increases to periods longer than a few days (Fig. 3b). The frequency of superflares on the slowly rotating stars (rotational period, >10 d) is only 1/20 of that of superflares on rapidly rotating stars. The rotation period correlates with the chromospheric activity, which is known to be an indicator of the magnetic activity of the stars¹⁷, and the more rapidly rotating stars have higher magnetic activity¹⁸. According to the dynamo theory of magnetic field generation, magnetic activity results from the interaction between rotation and convection¹⁹, and the rapid

rotation can cause the high magnetic activity. Our result implies that rapidly rotating stars with higher magnetic activity can cause more frequent superflares. The frequency distribution of superflares saturates for periods of less than a few days. A similar saturation is known for the relationship between the coronal X-ray activity and the rotation period²⁰.

The rotation period of a star is also known to be related to the star's age, with younger stars rotating more rapidly^{21,22}. Our findings suggest that superflares occur more frequently on the solar-type stars younger than the Sun. Moreover, on solar-type stars similar in age to the Sun, superflares occur less frequently but are nearly equal in energy to the superflares on the younger stars.

It has been pointed out that there is no record of solar superflares over the past 2,000 yr (ref. 3). According to the measurement of the impulsive nitrate events in polar ice, the largest proton flare event during the past 450 yr is the Carrington event²³, which occurred on 1 September 1859²⁴. The total energy released in this flare was estimated to be of order 10^{32} erg (ref. 25), which is only 1/1,000 of the maximum energy of flares on slowly rotating Sun-like stars. It has also been proposed that hot Jupiters have an important effect on stellar magnetic activity^{7,26,27} and that superflares occur only on solar-type stars with hot Jupiters. However, there is no hot Jupiter in the Solar System. For these reasons, it was suggested that a superflare on the Sun is extremely unlikely^{3,7}. Although the Kepler mission has discovered 1,235 exoplanet candidates around 997 host stars from a survey of 156,453 stars²⁸, no exoplanet has been found around the 148 G-type main-sequence stars with superflares. For a solar-type star with a hot Jupiter, the probability of a transit of the planet across the star is about 10% averaged over all possible orbital inclinations. If the superflares on all 148 stars were caused by hot Jupiters, then Kepler should detect 15 of them from transits. However, the Kepler planetary-transit search is almost complete for hot Jupiters²⁹, and the non-detection of planetary transits therefore suggests that hot Jupiters associated with superflares are rare.

Received 13 October 2011; accepted 14 March 2012.

Published online 16 May 2012.

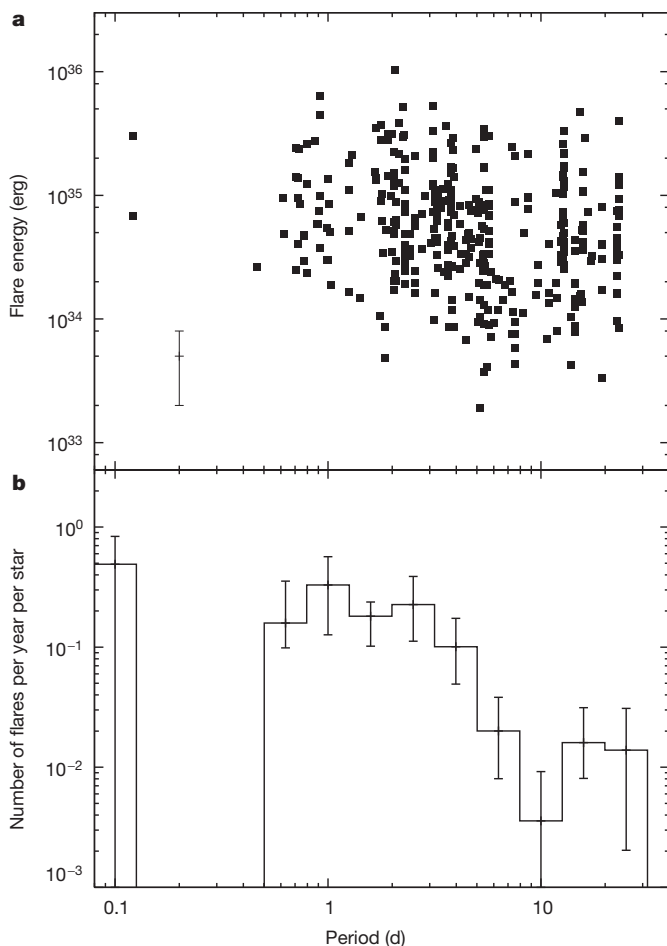


Figure 3 | Relations between the brightness variation period and the properties of flares on G-type main-sequence stars. **a**, Scatter plot of the flare energy as a function of the variation period. An apparent negative correlation between the variation period and the lower limit of the flare energy is caused by the detection limit of our flare-search method. The vertical bar in the panel indicates the typical uncertainty in the flare energy. **b**, Distribution of the occurrence of flares in each period bin as a function of variation period. The vertical axis indicates the number of flares with energy $\geq 5 \times 10^{34}$ erg per star per year. The error bars represent the 1σ uncertainty estimated from the uncertainty in the energy estimation and the square root of the event number in each period bin.

1. Baker, D. N. in *Space Weather: The Physics Behind a Slogan* (eds Scherer, K., Fichtner, H., Heber, B. & Mall, U.) 3–20 (Lect. Notes Phys. 656, Springer, 2004).
2. Schaefer, B. E. Flashes from normal stars. *Astrophys. J.* **337**, 927–933 (1989).
3. Schaefer, B. E., King, J. R. & Deliyannis, C. P. Superflares on ordinary solar-type stars. *Astrophys. J.* **529**, 1026–1030 (2000).
4. Gershberg, R. E. *Solar-Type Activity in Main-Sequence Stars* 191–380 (Springer, 2005).
5. Shibata, K. & Yokoyama, T. A Hertzsprung-Russell-like diagram for solar/stellar flares and corona: emission measure versus temperature diagram. *Astrophys. J.* **577**, 422–432 (2002).
6. Landini, M., Fossi, B. C., Pallavicini, R. & Piro, L. EXOSAT detection of an X-ray flare from the solar type star π^1 UMa. *Astron. Astrophys.* **157**, 217–222 (1986).
7. Rubenstein, E. P. & Schaefer, B. E. Are superflares on solar analogues caused by extrasolar planets? *Astrophys. J.* **529**, 1031–1033 (2000).
8. Koch, D. G. *et al.* Kepler mission design, realized photometric performance, and early science. *Astrophys. J.* **713**, L79–L86 (2010).
9. Brown, T. M., Latham, D. W., Everett, M. E. & Esquerdo, G. A. Kepler input catalog: photometric calibration and stellar classification. *Astron. J.* **142**, 112 (2011).
10. Kretschmar, M. The Sun as a star: observations of white-light flares. *Astron. Astrophys.* **530**, A84 (2011).
11. Rodono, M. *et al.* Rotational modulation and flares on RS CVn and BY Dra-type stars. I - Photometry and SPOT models for BY Dra, AU Mic, AR Lac, II Peg and V 711 Tau (= HR 1099). *Astron. Astrophys.* **165**, 135–156 (1986).
12. Kopal, Z. *Close Binary Systems* 147–261 (Chapman & Hall, 1959).
13. Unno, W., Osaki, Y., Ando, H., Saio, H. & Shibahashi, H. *Nonradial Oscillations of Stars* 237–255 (Univ. Tokyo Press, 1989).
14. Debusscher, J., Blomme, J., Aerts, C. & De Ridder, J. Global stellar variability study in the field-of-view of the Kepler satellite. *Astron. Astrophys.* **529**, A89 (2011).
15. Crosby, N. B., Aschwanden, M. J. & Dennis, B. R. Frequency distributions and correlations of solar X-ray flare parameters. *Sol. Phys.* **143**, 275–299 (1993).
16. Shakhovskaya, N. I. Stellar flare statistics - Physical consequences. *Sol. Phys.* **121**, 375–386 (1989).
17. Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K. & Vaughan, A. H. Rotation, convection, and magnetic activity in lower main-sequence stars. *Astrophys. J.* **279**, 763–777 (1984).
18. Pallavicini, R. *et al.* Relations among stellar X-ray emission observed from Einstein, stellar rotation and bolometric luminosity. *Astrophys. J.* **248**, 279–290 (1981).
19. Parker, E. N. *Cosmical Magnetic Fields: Their Origin and Their Activity* 532–749 (Oxford Univ. Press, 1979).

20. Randich, S. Coronal activity among open cluster stars. *ASP Conf. Ser.* **198**, 401–410 (2000).
21. Skumanich, A. Time scales for Call emission decay, rotational braking, and lithium depletion. *Astrophys. J.* **171**, 565–567 (1972).
22. Barnes, S. A. On the rotational evolution of solar- and late-type stars, its magnetic origins, and the possibility of stellar gyrochronology. *Astrophys. J.* **586**, 464–479 (2003).
23. Shea, M. A., Smart, D. F., McCracken, K. G., Dreschhoff, G. A. M. & Spence, H. E. Solar proton events for 450 years: the Carrington event in perspective. *Adv. Space Res.* **38**, 232–238 (2006).
24. Carrington, R. C. Description of a singular appearance seen in the Sun on September 1, 1859. *Mon. Not. R. Astron. Soc.* **20**, 13–15 (1859).
25. Tsurutani, B. T., Gonzalez, W. D., Lakhina, G. S. & Alex, S. The extreme magnetic storm of 1–2 September 1859. *J. Geophys. Res.* **108**, 1268–1275 (2003).
26. Cuntz, M., Saar, S. H. & Musielak, Z. E. On stellar activity enhancement due to interaction with extrasolar giant planets. *Astrophys. J.* **533**, L151–L154 (2000).
27. Ip, W.-H., Kopp, A. & Hu, J.-H. On the star-magnetosphere interaction of close-in exoplanets. *Astrophys. J.* **602**, L53–L56 (2004).
28. Borucki, W. J. *et al.* Characteristics of planetary candidates observed by Kepler. II. Analysis of the first four months of data. *Astrophys. J.* **736**, 19 (2011).
29. Howard, A. W. *et al.* Planet occurrence within 0.25 AU of solar-type stars from Kepler. Preprint at (<http://arxiv.org/abs/1103.2541>) (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Kepler was selected as NASA's tenth Discovery mission. Funding for the mission is provided by the NASA Science Mission Directorate. The data presented in this paper were obtained from the Multimission Archive at STScI. This work was supported by the Grant-in-Aid for the Global COE Program 'The Next Generation of Physics, Spun from Universality and Emergence' and the Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science and Technology of Japan. We are grateful to K. Sekiguchi and A. Hillier for suggestions.

Author Contributions H.M. was responsible for planning, coordination and data analysis. He also wrote the majority of the text. T.S. performed the data analysis for many stellar flares observed by Kepler, and S.N. and Y.N. analysed the rotation of stars observed by Kepler. T.N. and S.K. helped with data analysis and interpretation. S.H. and D.N. contributed to the interpretation of the stellar brightness variation and helped with the data analysis from a general stellar astronomical point of view. K.S. provided theoretical interpretation of the observations and gave advice on the paper's content.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.M. (maehara@kwasan.kyoto-u.ac.jp).

Tunable ion–photon entanglement in an optical cavity

A. Stute¹, B. Casabone¹, P. Schindler¹, T. Monz¹, P. O. Schmidt^{2,3}, B. Brandstätter¹, T. E. Northup¹ & R. Blatt^{1,4}

Proposed quantum networks require both a quantum interface between light and matter and the coherent control of quantum states^{1,2}. A quantum interface can be realized by entangling the state of a single photon with the state of an atomic or solid-state quantum memory, as demonstrated in recent experiments with trapped ions^{3,4}, neutral atoms^{5,6}, atomic ensembles^{7,8} and nitrogen-vacancy spins⁹. The entangling interaction couples an initial quantum memory state to two possible light–matter states, and the atomic level structure of the memory determines the available coupling paths. In previous work, the transition parameters of these paths determined the phase and amplitude of the final entangled state, unless the memory was initially prepared in a superposition state⁴ (a step that requires coherent control). Here we report fully tunable entanglement between a single ⁴⁰Ca⁺ ion and the polarization state of a single photon within an optical resonator. Our method, based on a bichromatic, cavity-mediated Raman transition, allows us to select two coupling paths and adjust their relative phase and amplitude. The cavity setting enables intrinsically deterministic, high-fidelity generation of any two-qubit entangled state. This approach is applicable to a broad range of candidate systems and thus is a promising method for distributing information within quantum networks.

Optical cavities are often proposed as a means to improve the efficiency of atom–photon entanglement generation. Experiments using single emitters^{3–5,9} collect photons over a limited solid angle, with only a small fraction of entanglement events detected. However, by placing the emitter inside a low-loss cavity, it is possible to generate photons with near-unit efficiency in the cavity mode^{1,10}. Neutral atoms in a resonator have been used to generate polarization-entangled photon pairs^{6,11}, but this has not yet been combined with coherent operations on the atomic state. Trapped ions have the advantage of well-developed methods for coherent state manipulation and readout^{12,13}. Using a single trapped ion integrated with a high-finesse cavity, we implement full tomography of the joint atom–photon state and generate maximally entangled states with fidelities up to $97.4 \pm 0.2\%$ (here the number in parentheses indicates the uncertainty in the last digit).

In initial demonstrations of atom–photon entanglement, the amplitudes of the resulting state are fixed by atomic transition amplitudes^{3,5,6,9,11}. If the final atomic states are not degenerate, as in the case of a Zeeman splitting, the phase of the atomic state after photon detection is determined by the time at which detection occurs. In contrast, we control both amplitude and phase via two simultaneous cavity-mediated Raman transitions. The bichromatic Raman fields ensure the independence of the atomic state from the photon-detection time; their relative amplitude and phase determine the state parameters. Within a quantum network, such a tunable state could be matched to any second state at a remote node, generating optimal long-distance entanglement in a quantum-repeater architecture¹⁴.

A tunable state has previously been used as the building block for teleportation⁴ and a heralded gate between remote qubits¹⁵. In these cases, tunability of the entangled state is inherited from control over

the initial state of the atom. The photonic qubit is encoded in frequency, and as a result, integration with a cavity would be technically challenging. The entangling process is intrinsically probabilistic, with efficiency limited to 50% even if all emitted photons could be collected. In the scheme presented here, the entangling interaction itself is tunable, and no coherent manipulation of the input state is required. For atomic systems with a complex level scheme in which several transition paths are possible, the two most suitable paths can be selected.

Our experimental apparatus (Fig. 1a) consists of a linear Paul trap storing a single ⁴⁰Ca⁺ ion within a 2 cm optical cavity^{16,17}. The cavity has a waist of 13 μm and finesse of 77,000 at 854 nm, the wavelength of the $4^2P_{3/2}$ – $3^2D_{5/2}$ transition. The rates of coherent atom–cavity coupling g , cavity-field decay κ , and atomic polarization decay γ are given by $(g, \kappa, \gamma) = 2\pi \times (1.4, 0.05, 11.2)$ MHz. The ion is located in both the waist and in an antinode of the cavity standing wave, and it is localized to within 13 ± 7 nm along the cavity axis¹⁷. Entanglement is generated via a bichromatic Raman field at 393 nm and read out using a quadrupole field at 729 nm.

A magnetic field of 2.96 G is applied along the quantization axis \hat{z} and perpendicular to the cavity axis. The cavity supports degenerate horizontal (H) and vertical (V) polarization modes, where H is defined parallel to \hat{z} . At the cavity output, the modes are separated on a polarizing beamsplitter and detected at avalanche photodiodes. A half- and a quarter-waveplate before the beamsplitter allow us to set the measurement basis of the photon¹⁸.

The entangling process is illustrated in Fig. 1b. Following a Doppler-cooling interval, the ion is initialized via optical pumping in the state $|S\rangle \equiv |4^2S_{1/2}, m_S = -1/2\rangle$. In order to couple $|S\rangle$ simultaneously to the two states $|D\rangle \equiv |3^2D_{5/2}, m_D = -3/2\rangle$ and $|D'\rangle \equiv |3^2D_{5/2}, m_D = -5/2\rangle$, we apply a phase-stable bichromatic Raman field, detuned by Δ_1 and Δ_2 from the $|S\rangle$ – $|P\rangle$ transition. Here, the intermediate state $|P\rangle \equiv |4^2P_{3/2}, m_P = -3/2\rangle$ is used. The cavity is stabilized at detuning $\Delta_1^c \approx -400$ MHz from the $|P\rangle$ – $|D\rangle$ transition and $\Delta_2^c = \Delta_1^c + \Delta_{D,D'}$ from the $|P\rangle$ – $|D'\rangle$ transition, where $\Delta_{D,D'}$ is the Zeeman splitting between $|D\rangle$ and $|D'\rangle$. When Δ_i^c and Δ_i satisfy the Raman resonance condition for both $i = (1, 2)$, population is transferred coherently from $|S\rangle$ to both $|D\rangle$ and $|D'\rangle$, and a single photon is generated in the cavity^{19–22}.

The effective coupling strength of each of the two transitions is given by $g_i^{\text{eff}} = \Omega_i G_i g / \Delta_i$. Here, Ω_1 and Ω_2 are the amplitudes of the Raman fields, and G_1 and G_2 are the products of the Clebsch-Gordon coefficients and the projections of laser and vacuum-mode polarizations onto the atomic dipole moment¹⁷. In free space, these two pathways generate π - and σ^+ -polarized photons, respectively. Within the cavity, the π photon is projected onto H and the σ^+ photon onto V ^{16,17}. Ideally, the bichromatic fields generate any state of the form

$$|\psi\rangle = \cos \alpha |DH\rangle + e^{i\varphi} \sin \alpha |D'V\rangle$$

where $\alpha \equiv \tan^{-1}(g_2^{\text{eff}}/g_1^{\text{eff}})$ and φ is determined by the relative phase of the Raman fields. To determine the overlap of the measured state with $|\psi\rangle$, we perform quantum state tomography of the ion–photon density matrix ρ for given values of α and φ . Ion and photon are measured in all

¹Institut für Experimentalphysik, Universität Innsbruck, Technikerstraße 25, 6020 Innsbruck, Austria. ²QUEST Institute for Experimental Quantum Metrology, Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany. ³Institut für Quantenoptik, Leibniz Universität Hannover, 30167 Hannover, Germany. ⁴Institut für Quantenoptik und Quanteninformation der Österreichischen Akademie der Wissenschaften, Technikerstraße 21a, 6020 Innsbruck, Austria.

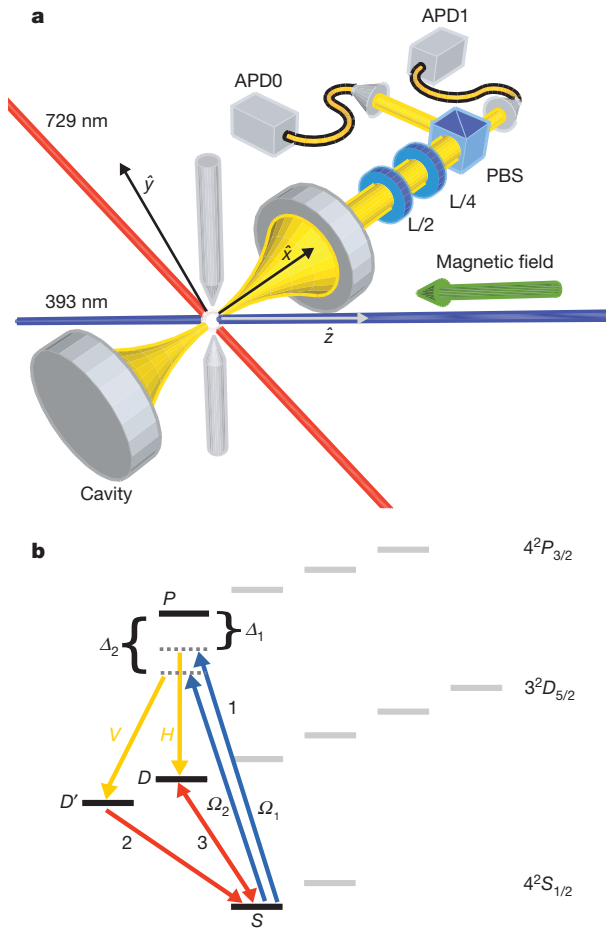


Figure 1 | Experimental apparatus and entanglement sequence. **a**, An ion is confined in a Paul trap (indicated by two endcaps) at the point of maximum coupling to a high-finesse cavity. A 393-nm laser (blue) generates atom–photon entanglement, characterized using a 729-nm laser (red). The polarization of photons leaving the cavity is analysed using half- and quarter-waveplates (L/2, L/4), a polarizing beamsplitter cube (PBS), and fibre-coupled avalanche photodiodes (APD0, APD1). **b**, A bichromatic Raman pulse (blue arrows labelled 1) with Rabi frequencies Ω_1 , Ω_2 and detunings Δ_1 , Δ_2 couples $|S\rangle$ to states $|D\rangle$ and $|D'\rangle$ via two cavity modes H and V (yellow arrows), generating a single cavity photon. To read out entanglement, $|D'\rangle$ is mapped to $|S\rangle$ (red arrow 2), and coherent operations on the S – D transition (red arrow 3) prepare the ion for measurement. Solid black levels indicate the four relevant states $|S\rangle$, $|D\rangle$, $|D'\rangle$ and $|P\rangle$. Solid grey levels indicate all additional states in the $4^2S_{1/2}$, $3^2D_{5/2}$ and $4^2P_{3/2}$ manifolds of $^{40}\text{Ca}^+$. Dashed grey levels indicate the Raman resonance frequencies.

nine combinations of ion Pauli bases $\{\sigma_x, \sigma_y, \sigma_z\}$ and photon polarization bases $\{H/V, \text{diagonal/antidiagonal, right/left}\}^{18}$.

In order to measure the ion in all three bases, we first map the superposition of $\{|D', D\rangle\}$ onto the $\{|S, D\rangle\}$ states^{12,13}. We then perform additional coherent operations to select the measurement basis and discriminate between S and D via fluorescence detection¹³. Each sequence lasts 1.5 ms and consists of 800 μs of Doppler cooling, 60 μs of optical pumping, a 40 μs Raman pulse, a 4 μs mapping pulse, an optional 4.3 μs rotation, and 500 μs of fluorescence detection. The probability of detecting a photon in a single sequence is 5.7%; we thus detect on average 40.5 events s^{-1} . Note that the photon is generated with near-unit efficiency, and detection is primarily limited by the probability of the photon leaving the cavity (16%) and the photodiode efficiencies (40%).

In a first set of measurements, we choose the case $\alpha = \pi/4$, corresponding to a maximally entangled state $|\psi\rangle$. From the tomographic data, the density matrix is reconstructed as shown in Fig. 2a. Here we have tuned Ω_1 and Ω_2 so as to produce both photon polarizations with

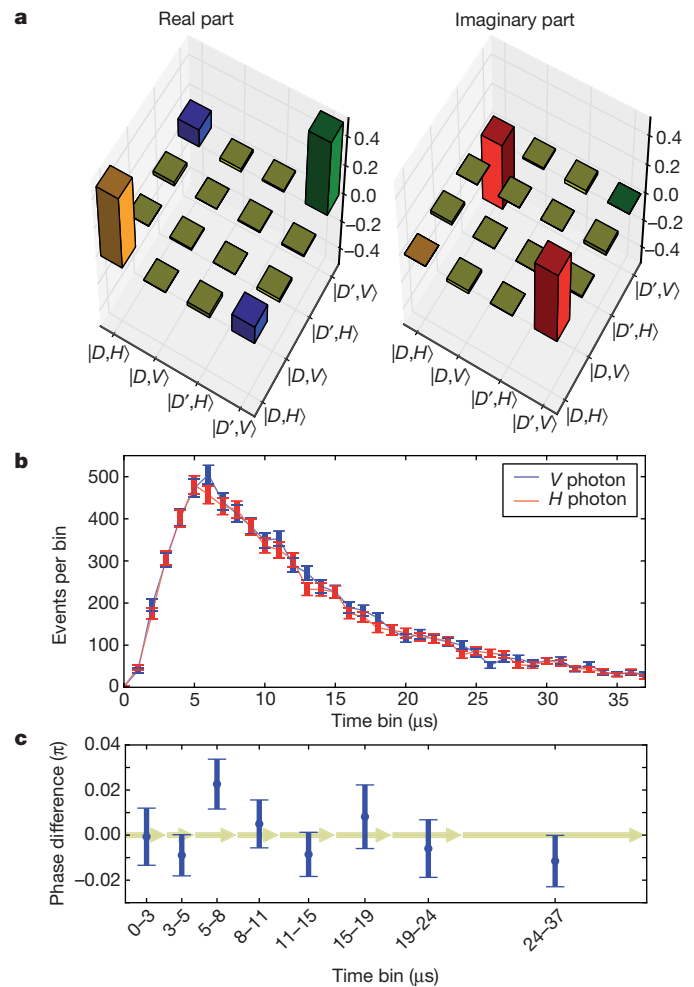


Figure 2 | Quantum state tomography of the joint ion–photon state, containing $\sim 40,000$ events. **a**, Real and imaginary parts of all density matrix elements for Raman phase $\varphi = 0.25$, from which a fidelity $F = 97.4 \pm 0.2\%$ is calculated. Colours for the density matrix elements correspond to those used in Figs 3a and 4a. **b**, Temporal pulse shape of H and V cavity photons. Error bars represent one s.d., based on Poissonian photon statistics. **c**, Phase of the ion–photon state versus photon-detection time. Arrows indicate time-bin intervals of the tomography data. Error bars represent one s.d. (see Methods).

equal probability, corresponding to maximal overlap of the temporal pulse shapes of H and V photons (Fig. 2b). In order to demonstrate that the photon-detection time does not determine the phase of the state, we extract this phase from state tomography as a function of the photon time bin (Fig. 2c). Because the frequency difference of the bichromatic fields $\Delta_1 - \Delta_2$ is equal to the level spacing between $|D\rangle$ and $|D'\rangle$, the phase $\varphi = 0.25\pi$ remains constant. Further details are given in the Methods section.

Tomography over all time bins yields a fidelity of $F \equiv \langle \psi | \rho | \psi \rangle = 97.4 \pm 0.2\%$ with respect to the maximally entangled state, placing our system definitively in the non-classical regime $F > 50\%$. Another two-qubit entanglement witness is the concurrence²³, which we calculate to be $95.2 \pm 0.5\%$. The observed entanglement can also be used to test local hidden-variable models (LHVMs) via the violation of the Clauser–Horne–Shimony–Holt (CHSH) Bell inequality²⁴. Entanglement of a hybrid atom–photon system holds particular interest, because it could be used for a loophole-free test of a Bell-type inequality²⁵. Whereas LHVMs require the Bell observable of the CHSH inequality to be less than 2, we measure a value $2.75 \pm 0.01\%$ > 2 , where quantum mechanics provides an upper bound of $2\sqrt{2}$.

We now establish that we can prepare $|\psi\rangle$ with high fidelity over the full range of the Raman phase φ . We repeat state tomography for seven

additional values of the relative Raman phase. As a function of φ , the real and imaginary parts of the coherence $\rho_{14} \equiv \langle DH|\rho|D'V \rangle$ vary sinusoidally as expected (Fig. 3a). The fidelity has a mean value of $96.9 \pm 0.1\%$ and does not vary within error bars over all target phases (Fig. 3c).

A second measurement set demonstrates control over the amplitudes $\cos \alpha$ and $\sin \alpha$ of the entangled ion–photon state. After selecting three target amplitudes $\cos \alpha = \{1/\sqrt{2}, 1/\sqrt{3}, 1/\sqrt{8}\}$, we generate each corresponding state by adjusting the Raman field amplitudes, since α is a function of the ratio Ω_2/Ω_1 . The density matrix for each state is then measured. In Fig. 4a, we see that the populations $\rho_{11} \equiv \langle DH|\rho|DH \rangle$ and $\rho_{44} \equiv \langle D'V|\rho|D'V \rangle$ for the three target amplitudes agree well with theoretical values. The fidelities of the asymmetric states (Fig. 4b) are as high as those of the maximally entangled states and are limited by the populations, that is, by errors in tuning the Raman fields to match the target values.

Errors in atomic state detection^{5,25}, atomic decoherence¹¹ and multiple excitations of the atom³ reduce the fidelity of the atom–photon entangled state by $\ll 1\%$. Imperfect initialization and manipulation of the ion due to its finite temperature and laser intensity fluctuations decrease the fidelity by 1%. The two most significant reductions in fidelity are due to dark counts of the avalanche photodiodes at a rate of 36 Hz (1.5%) and imperfect overlap of the temporal pulse shapes (1%).

To our knowledge, the measurements reported above represent both the highest fidelity and the fastest rate of entanglement detection to date between a photon and a single-emitter quantum memory. This detection rate is limited by the fact that most cavity photons are absorbed or scattered by the mirror coatings, and only 16% enter the output mode. However, using mirrors with state-of-the-art losses and a highly asymmetric transmission ratio, an output coupling efficiency exceeding 99% is possible (see Methods). In contrast, without a cavity, using a lens of numerical aperture 0.5 to collect photons, the efficiency would be 6.7%. In addition, the infrared wavelength of the output photons is well-suited to fibre distribution, enabling long distance quantum networks. We note that a faster detection rate could be achieved by triggering ion-state readout on the detection of a photon.

We have demonstrated full control of the phase and amplitude of an entangled ion–photon state, which opens up new possibilities for

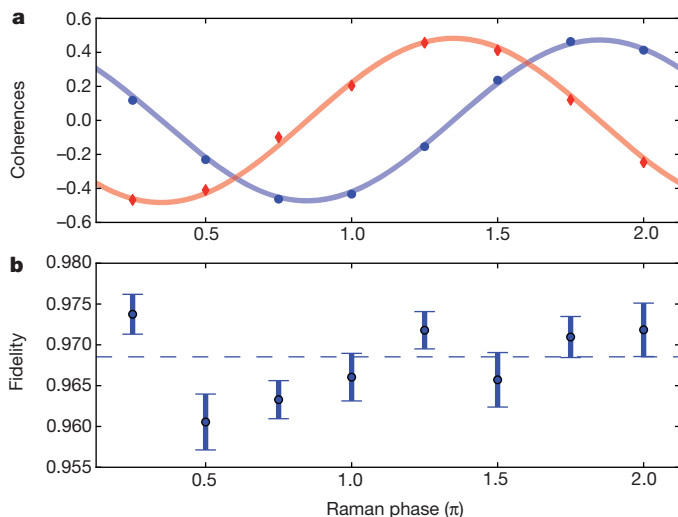


Figure 3 | State tomography as a function of Raman phase ($\sim 340,000$ events). **a**, Ion–photon state coherences, defined as $\text{Re}(\rho_{14})$ (filled blue circles) and $\text{Im}(\rho_{14})$ (filled red diamonds) as a function of Raman phase. Error bars are smaller than the size of the symbols. Each value is extracted from a full state tomography of ρ as in Fig. 2a. Both curves are fitted simultaneously, with phase offset constrained to $\pi/2$. The fit contrast is $95.6 \pm 0.4\%$. **b**, Fidelities of the eight states, with a dashed line indicating the mean value. Error bars represent one s.d. (see Methods).

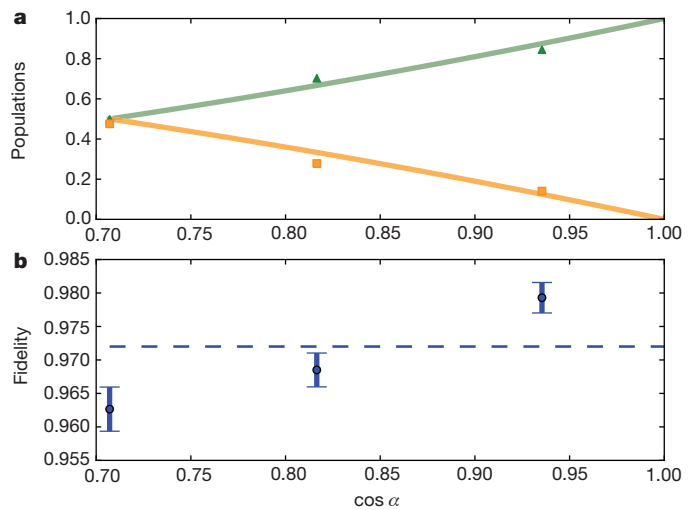


Figure 4 | State tomography for three values of amplitude, $\cos \alpha$. **a**, Ion–photon state populations, defined as the density matrix elements ρ_{11} (orange squares) and ρ_{44} (green triangles) are plotted for the three target amplitudes $\cos \alpha = \{1/\sqrt{2}, 1/\sqrt{3}, 1/\sqrt{8}\}$. Error bars are smaller than the size of the symbols. Solid lines represent the amplitudes of the target states. **b**, The corresponding fidelities are $F = \{96.3 \pm 0.3, 96.8 \pm 0.3, 98.0 \pm 0.4\}\%$. A dashed line indicates the mean value. Error bars represent one s.d. (see Methods).

quantum communication schemes. In contrast to monochromatic schemes, evolution of the relative phase of the atomic state after photon detection is determined only by the start time of the experiment and not by the photon-detection time. The state $|\psi\rangle$ is in this sense predetermined and can be stored in, or extracted from, a quantum memory in a time-independent manner. The bichromatic Raman process used here provides a basis for coherent atom–photon state mapping as well as one- or two-dimensional cluster state generation²⁶.

METHODS SUMMARY

Detection and state tomography. The cavity output path branches at a polarizing beamsplitter into two measurement paths, and the detection efficiencies of these paths are unequal. We compensate for this imbalance by performing two measurements for a given choice of ion and photon basis and sum the results; between the measurements, a rotation of the output waveplates swaps the two paths.

Correlations of the photon polarization and the atomic state are the input for maximum likelihood reconstruction of the most likely states²⁷. Error bars are one standard deviation derived from non-parametric bootstrapping²⁸ assuming a multinomial distribution.

Time independence. In the experiments of refs 3, 6 and 9, although the phase of the entangled state is time independent before photon detection, the phase of the atomic state after photon detection evolves due to Larmor precession. It is thus necessary to fix the time between photon detection and atomic state readout in order to measure the same φ for all realizations of the experiment. In contrast, for the case of Raman fields $\Omega_1 e^{i\omega_1 t}$ and $\Omega_2 e^{i\omega_2 t}$, the correct choice of frequency $\omega_1 - \omega_2 = \omega_D - \omega_C$ means that both the phase of the entangled atom–photon state before photon detection and the phase of the atomic state after photon detection are independent of photon detection time.

We define a model system with bases $\{|S, n\rangle, |D, n\rangle, |D', n\rangle\}$, where $n = \{0, 1\}$ is the photon number in either of the two degenerate cavity modes. The excited state has been adiabatically eliminated, so that g_1^{eff} couples $|S, 0\rangle$ to $|D, 1\rangle$ and g_2^{eff} couples $|S, 0\rangle$ to $|D', 1\rangle$. After transformation into a rotating frame $U = e^{i\omega_1 t|S\rangle\langle S|} e^{i(\omega_1 - \omega_2)t|D'\rangle\langle D'|}$, the Hamiltonian is

$$(\omega_S - \omega_1)|S\rangle\langle S| + (\omega_D - \omega_C)|D\rangle\langle D| + (\omega_{D'} - (\omega_1 - \omega_2))|D'\rangle\langle D'| + \omega_C|1\rangle\langle 1| + (g_1^{\text{eff}}|D, 1\rangle\langle S, 0| + g_2^{\text{eff}}|D', 1\rangle\langle S, 0| + \text{h.c.}),$$

where $\hbar = 1$, $\{\omega_S, \omega_D, \omega_{D'}\}$ are the state frequencies, ω_C is the cavity frequency, and terms rotating at $|\omega_1 - \omega_2| \gg g_i^{\text{eff}}$ are omitted²⁹. In this frame, the couplings g_i^{eff} are time-independent, and the states $|D\rangle$ and $|D'\rangle$ are degenerate. Therefore, the phase between $|D, 1\rangle$ and $|D', 1\rangle$ remains fixed during Raman transfer, and the phase between $|D, 0\rangle$ and $|D', 0\rangle$ stays constant after photon detection.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 February; accepted 11 April 2012.

1. Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
2. Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
3. Blinov, B. B., Moehring, D. L., Duan, L. M. & Monroe, C. Observation of entanglement between a single trapped atom and a single photon. *Nature* **428**, 153–157 (2004).
4. Olmschenk, S. *et al.* Quantum teleportation between distant matter qubits. *Science* **323**, 486–489 (2009).
5. Volz, J. *et al.* Observation of entanglement of a single photon with a trapped atom. *Phys. Rev. Lett.* **96**, 030404 (2006).
6. Wilk, T., Webster, S. C., Kuhn, A. & Rempe, G. Single-atom single-photon quantum interface. *Science* **317**, 488–490 (2007).
7. Matsukevich, D. N. *et al.* Entanglement of a photon and a collective atomic excitation. *Phys. Rev. Lett.* **95**, 040405 (2005).
8. Sherson, J. *et al.* Quantum teleportation between light and matter. *Nature* **443**, 557–560 (2006).
9. Togan, E. *et al.* Quantum entanglement between an optical photon and a solid-state spin qubit. *Nature* **466**, 730–734 (2010).
10. Law, C. & Kimble, H. Deterministic generation of a bit-stream of single-photon pulses. *J. Mod. Opt.* **44**, 2067–2074 (1997).
11. Weber, B. *et al.* Photon-photon entanglement with a single trapped atom. *Phys. Rev. Lett.* **102**, 030501 (2009).
12. Leibfried, D., Blatt, R., Monroe, C. & Wineland, D. Quantum dynamics of single trapped ions. *Rev. Mod. Phys.* **75**, 281–324 (2003).
13. Häffner, H., Roos, C. & Blatt, R. Quantum computing with trapped ions. *Phys. Rep.* **469**, 155–203 (2008).
14. Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
15. Maunz, P. *et al.* Heralded quantum gate between remote quantum memories. *Phys. Rev. Lett.* **102**, 250502 (2009).
16. Russo, C. *et al.* Raman spectroscopy of a single ion coupled to a high-finesse cavity. *Appl. Phys. B* **95**, 205–212 (2009).
17. Stute, A. *et al.* Toward an ion-photon quantum interface in an optical cavity. *Appl. Phys. B* <http://dx.doi.org/10.1007/s00340-011-4861-0> (published online, 13 January 2012).
18. James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001).
19. McKeever, J. *et al.* Deterministic generation of single photons from one atom trapped in a cavity. *Science* **303**, 1992–1994 (2004).
20. Keller, M., Lange, B., Hayasaka, K., Lange, W. & Walther, H. Continuous generation of single photons with controlled waveform in an ion-trap cavity system. *Nature* **431**, 1075–1078 (2004).
21. Hijkema, M. *et al.* A single-photon server with just one atom. *Nature Phys.* **3**, 253–255 (2007).
22. Barros, H. G. *et al.* Deterministic single-photon source from a single ion. *N. J. Phys.* **11**, 103004 (2009).
23. Wootters, W. K. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.* **80**, 2245–2248 (1998).
24. Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
25. Rosenfeld, W. *et al.* Towards a loophole-free test of Bell's inequality with entangled pairs of neutral atoms. *Adv. Sci. Lett.* **2**, 469–474 (2009).
26. Economou, S. E., Lindner, N. & Rudolph, T. Optically generated 2-dimensional photonic cluster state from coupled quantum dots. *Phys. Rev. Lett.* **105**, 093601 (2010).
27. Ježek, M., Fiurášek, J. & Hradil, Z. Quantum inference of states and processes. *Phys. Rev. A* **68**, 012305 (2003).
28. Davison, A. & Hinkley, D. *Bootstrap Methods and Their Application* (Cambridge Univ. Press, 1997).
29. Shore, B. *The Theory of Coherent Atomic Excitation* (Wiley, 1990).

Acknowledgements We thank J. Barreiro, D. Nigg, K. Hammerer and W. Rosenfeld for discussions. This work was supported by the Austrian Science Fund (FWF), the European Commission (AQUTE), the Institut für Quanteninformation GmbH, and a Marie Curie International Incoming Fellowship within the 7th European Framework Program.

Author Contributions Experiments were performed by A.S., B.C. and T.E.N., with contributions from P.S. to the set-up. Data analysis was performed by A.S., B.C. and T.M. The experiment was conceived by P.O.S. and R.B. and further developed in discussions with A.S., B.B., B.C. and T.E.N. All authors contributed to the discussion of results and participated in manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.E.N. (tracy.northup@uibk.ac.at).

METHODS

Detection and state tomography. The cavity output path branches at a polarizing beamsplitter into two measurement paths, and the detection efficiencies of these paths are unequal. We compensate for this imbalance by performing two measurements for a given choice of ion and photon basis and sum the results; between the measurements, a rotation of the output waveplates swaps the two paths.

At each measurement setting, we record on average 4,722 events in which a single photon has been detected. While a photon is detected in 5.7% of sequences, the atom is always measured. Correlations of the photon polarization and the atomic state are the input for maximum likelihood reconstruction of the most likely states²⁷. Error bars are one standard deviation derived from non-parametric bootstrapping²⁸ assuming a multinomial distribution.

Time independence. The phase of the entangled atom–photon state is inferred from the measurements of photon polarization and atomic-state phase. In the experiments of refs 3, 6 and 9, although the phase of the entangled state is time independent before photon detection, the phase of the atomic state after photon detection evolves due to Larmor precession. It is thus necessary to fix the time between photon detection and atomic state readout in order to measure the same φ for all realizations of the experiment. In contrast, for the case of Raman fields $\Omega_1 e^{i\omega_1 t}$ and $\Omega_2 e^{i\omega_2 t}$, the correct choice of frequency $\omega_1 - \omega_2 = \omega_{D'} - \omega_D$ means that both the phase of the entangled atom–photon state before photon detection and the phase of the atomic state after photon detection are independent of photon-detection time.

We define a model system with bases $\{|S, n\rangle, |D, n\rangle, |D', n\rangle\}$, where $n = \{0, 1\}$ is the photon number in either of the two degenerate cavity modes. The excited state

has been adiabatically eliminated, so that g_1^{eff} couples $|S, 0\rangle$ to $|D, 1\rangle$ and g_2^{eff} couples $|S, 0\rangle$ to $|D', 1\rangle$. After transformation into a rotating frame $U = e^{i\omega_1 t}|S\rangle\langle S|e^{i(\omega_1 - \omega_2)t}|D'\rangle\langle D'|}$, the Hamiltonian is

$$(\omega_S - \omega_1)|S\rangle\langle S| + \omega_D|D\rangle\langle D| + (\omega_{D'} - (\omega_1 - \omega_2))|D'\rangle\langle D'| + \omega_C|1\rangle\langle 1| \\ + (g_1^{\text{eff}}|D, 1\rangle\langle S, 0| + g_2^{\text{eff}}|D', 1\rangle\langle S, 0| + \text{h.c.}),$$

where $\hbar = 1$, $\{\omega_S, \omega_D, \omega_{D'}\}$ are the state frequencies, ω_C is the cavity frequency, and terms rotating at $|\omega_1 - \omega_2| \gg g_i^{\text{eff}}$ are omitted²⁹. In this frame, the couplings g_i^{eff} are time-independent, and the states $|D\rangle$ and $|D'\rangle$ are degenerate. Therefore, the phase between $|D, 1\rangle$ and $|D', 1\rangle$ remains fixed during Raman transfer, and the phase between $|D, 0\rangle$ and $|D', 0\rangle$ stays constant after photon detection.

Cavity parameters. The cavity mirrors have transmission $T_1 = 13$ p.p.m. and $T_2 = 1.3$ p.p.m., with combined losses of 68 p.p.m. State-of-the-art combined losses at this wavelength are $L = 4$ p.p.m. (ref. 30). In our cavity, these losses would correspond to an output coupling efficiency of $T_1/(T_1 + T_2 + L) = 71\%$. To improve this efficiency, an output mirror with higher transmission T_1 could be used; for example, $T_1 = 500$ p.p.m. corresponds to an efficiency of 99%. The cavity decay rate κ would also increase, but single-photon generation with near-unit efficiency is valid in the bad-cavity regime¹⁰.

30. Rempe, G., Thompson, R. J., Brecha, R. J., Lee, W. D. & Kimble, H. J. Optical bistability and photon statistics in cavity quantum electrodynamics. *Phys. Rev. Lett.* **67**, 1727–1730 (1991).

All-solid-state dye-sensitized solar cells with high efficiency

In Chung¹, Byunghong Lee², Jiaqing He¹, Robert P. H. Chang² & Mercouri G. Kanatzidis¹

Dye-sensitized solar cells based on titanium dioxide (TiO₂) are promising low-cost alternatives to conventional solid-state photovoltaic devices based on materials such as Si, CdTe and CuIn_{1-x}Ga_xSe₂ (refs 1, 2). Despite offering relatively high conversion efficiencies for solar energy, typical dye-sensitized solar cells suffer from durability problems that result from their use of organic liquid electrolytes containing the iodide/tri-iodide redox couple, which causes serious problems such as electrode corrosion and electrolyte leakage³. Replacements for iodine-based liquid electrolytes have been extensively studied, but the efficiencies of the resulting devices remain low³⁻⁹. Here we show that the solution-processable p-type direct bandgap semiconductor CsSnI₃ can be used for hole conduction in lieu of a liquid electrolyte. The resulting solid-state dye-sensitized solar cells consist of CsSnI_{2.95}F_{0.05} doped with SnF₂, nanoporous TiO₂ and the dye N719, and show conversion efficiencies of up to 10.2 per cent (8.51 per cent with a mask). With a bandgap of 1.3 electronvolts, CsSnI₃ enhances visible light absorption on the red side of the spectrum to outperform the typical dye-sensitized solar cells in this spectral region.

Photovoltaics is a promising renewable energy technology that converts sunlight to electricity, with broad potential to contribute significantly to solving the future energy problem that humanity faces. To date, semiconductor solar cells dominate commercial markets, with crystalline Si having an 80% share; the remaining 20% is mostly thin-film solar technology, such as CdTe and CuIn_{1-x}Ga_xSe₂ (ref. 10). The former is an indirect bandgap semiconductor typically requiring a 300-μm-thick absorption layer, and material and processing costs are very high. The latter contains elements that are toxic and of low abundance in the Earth. However CuIn_{1-x}Ga_xSe₂ forms the best performing thin-film solar devices, exhibiting an efficiency of ~20%, but is more than 1.4 times as expensive as CdTe and amorphous Si. A low-cost and environmentally friendly alternative to these solid-state devices is the dye-sensitized solar cell (DSC)^{1,2}. It is inexpensive to prepare, and the light-weight thin-film structures are compatible with automated manufacturing.

Conventional DSCs (Grätzel cells) consist of a self-assembled monolayer of molecular dye at the interface between a mesoporous wide-bandgap semiconductor oxide and a liquid electrolyte^{1,2}. The most commonly used redox couple is iodide/tri-iodide (I⁻/I₃⁻) in an organic liquid electrolyte—however, it is highly corrosive, volatile and photoreactive, interacting with common metallic components and sealing materials. Consequently, it adversely affects long-term performance and durability³. Alternative liquid electrolytes free of the I⁻/I₃⁻ redox couple have been a long-term goal in this field^{5,6,11}. The ultimate solutions would be purely solid-state cells, given the inevitable problems of any liquid electrolyte, such as leakage, heavy weight and complex chemistry. Efforts have focused on using solid-state organic or p-type conducting polymer hole-transport materials (HTMs), but their conversion efficiency remains modest. Solar cells using spiro-OMeTAD (refs 9, 12) and bis-EDOT (ref. 7) exhibit the highest conversion efficiencies among organic and conducting polymer materials of 6.08% and 6.1%, respectively. (TAD is

2,2',7,7'-tetrakis(*N,N*-di-*p*-methoxyphenyl-amine)9,9'-spirobifluorene, and EDOT is 2,2'-bis(3,4-ethylenedioxythiophene).) A general problem of solid HTMs in DSCs is poor filling of the nanoporous TiO₂ layer; this interrupts the hole-conducting path between the HTM and the dye molecule adsorbed on TiO₂ (ref. 1). Despite many anticipated advantages, inorganic HTMs are uncommon. CuI, CuSCN and NiO are examples, but their mobilities are very low. The efficiency of CuI-based DSCs was found to initially reach about 3%, but rapidly photodegraded¹. Cells of CuSCN (ref. 13) and p-type NiO particles¹⁴ showed low efficiencies.

Here we report a new type of all-solid-state, inorganic solar cell system that consists of the p-type direct bandgap semiconductor CsSnI₃ and n-type nanoporous TiO₂ with the dye N719 (*cis*-diisothiocyanato-bis(2,2'-bipyridyl-4,4'-dicarboxylato) ruthenium(II) bis-(tetrabutylammonium)). We show that CsSnI₃ is well fitted for this purpose because of its energy gap of 1.3 eV and a remarkably high hole mobility of $\mu_h = 585 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ at room temperature. We found that CsSnI₃ is soluble in polar organic solvents, such as acetonitrile, *N,N*-dimethylformamide and methoxyacetonitrile. Consequently, it is solution-processable and can be transferred into TiO₂ pores at a molecular level to make intimate contacts with dye molecules and TiO₂. We present results showing that doping of CsSnI₃ with F and

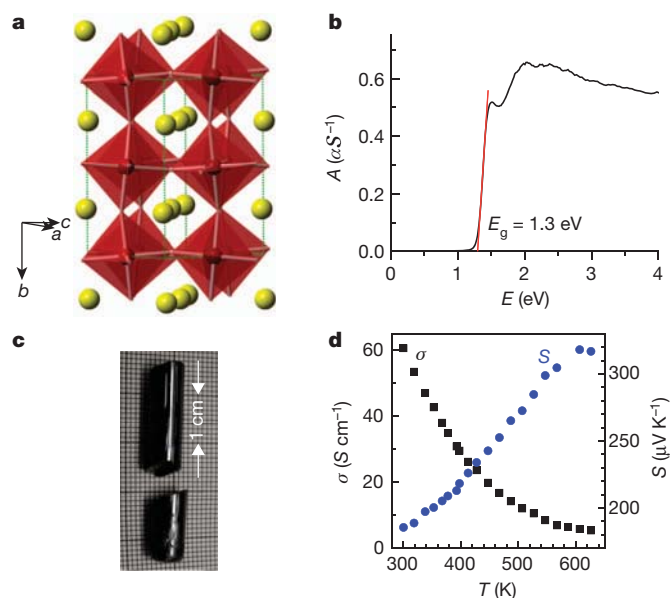


Figure 1 | Crystal structure and optical and electrical transport properties of CsSnI₃. **a**, Distorted three-dimensional perovskite structure of CsSnI₃ at room temperature. Red polyhedron, [SnI_{6/2}]⁻; yellow sphere, Cs. **b**, Sharp absorption edge at 1.3 eV of CsSnI₃. **A**, absorption in units of αS^{-1} , where α is the absorption coefficient and S the scattering coefficient. **E**, energy in eV. E_g , the value of the energy gap. **c**, A typical ingot of CsSnI₃ grown in a Bridgman furnace. **d**, Temperature dependence of electrical conductivity (σ , filled squares) and Seebeck coefficient (S , filled circles).

¹Department of Chemistry, Northwestern University, Evanston, Illinois 60208, USA. ²Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208 USA.

SnF_2 dramatically improves the photocurrent density (J_{sc}) and power conversion efficiency (η). At an optimum molar concentration of 5% F and 5% SnF_2 , the cell exhibits the highest efficiency so far reported for a solid-state solar cell equipped with a dye-sensitizer: $\eta = 10.2\%$ under the standard air mass 1.5 (AM 1.5) irradiation (100 mW cm^{-2}), and $\eta = 8.51\%$ with a mask. The observed value is close to that of the highest reported performance N719-dye-containing Grätzel cell ($\eta \approx 11\%$)¹⁵.

In our cells, CsSnI_3 replaces the entire liquid electrolyte. Optical absorption spectroscopy and incident photon-to-current conversion efficiency (IPCE) measurements show that our cells exhibit a red-shifted absorption edge (at 789 nm) compared to that of the typical Grätzel cell and outperform it in the red and infrared spectral regions. We note that poor performance of DSCs in this spectral range has been one of their main drawbacks.

CsSnI_3 adopts a distorted three-dimensional perovskite structure that crystallizes in the orthorhombic $Pnma$ space group at room temperature (Fig. 1a)¹⁶. The compound is a direct bandgap semiconductor with a sharp optical absorption edge at 1.3 eV (Fig. 1b). Hall effect measurements on a CsSnI_3 thin film showed p-type conduction behaviour and a very high hole mobility of $\mu_{\text{h}} = 585 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ at room temperature. To confirm this, we prepared large, bubble-free and crack-free polycrystalline ingots (Fig. 1c), and examined the electrical properties of the ingot. Thermoelectric power measurements gave positive Seebeck coefficients over the entire temperature range with linear dependence on temperature, suggesting p-type conduction (Fig. 1d). These hole mobilities are two to three orders of magnitude larger than those of any organic polymer HTM.

A schematic diagram of the relative energy levels of CsSnI_3 , TiO_2 and the N719 dye is shown in Fig. 2. The positions of the valence band maximum (VBM) and the conduction band minimum (CBM) of CsSnI_3 were determined by the work function (4.92 eV) according to ultraviolet photoemission spectroscopy and the energy gap (1.3 eV). The energy levels of TiO_2 and the N719 dye are taken from the literature^{5,17}. The diagram validates the excellent fit of CsSnI_3 in our new solid-state solar cell system, giving perfect charge separation and replacing liquid electrolytes. The CBM lies nearly in the same energy level as the lowest unoccupied molecular orbital (LUMO) of the N719 dye and above the CBM of TiO_2 . The VBM lies much higher than the highest occupied molecular orbital (HOMO) of the dye. From a conventional view of DSCs, electrons generated by the dye are transported to n-type

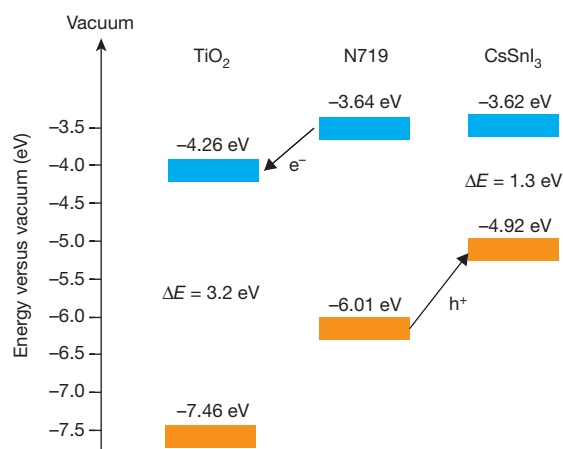


Figure 2 | Energy levels of the components of the CsSnI_3 solid-state solar cell. The valence band maximum (VBM; orange colour) and the conduction band minimum (CBM; blue colour) of TiO_2 and CsSnI_3 are represented in eV, along with the energy difference between the edges. The ground (HOMO; orange) and excited states (LUMO; blue) of N719 dye is also shown. The energy scale is referenced to the vacuum level. ΔE , e^- and h^+ represent the value of the energy gap, electron and hole, respectively.

semiconductor TiO_2 and the oxidized dye is readily regenerated by p-type semiconductor CsSnI_3 because it is a fast hole transporter.

The cells we report here are likely to operate by a mechanism that is different from that of conventional DSCs incorporating liquid electrolyte. For example, solid-state cells made with $\text{CsSnI}_{3-x}\text{F}_x$ and TiO_2 without the dye worked as a photovoltaic devices but with a low efficiency, $\sim 0.2\%$ (Supplementary Information). This demonstrates that $\text{CsSnI}_{3-x}\text{F}_x$ can act as its own dye and photogenerated holes and electrons can be separated, but it is not very efficient by itself. So, if in our dye-containing cells CsSnI_3 absorbs the light instead of the dye, electrons and holes would be generated therein. Electrons would then be transferred to the CBM of TiO_2 . At this stage, the N719 dye could act a conduit for fast charge transfer from CsSnI_3 facilitating rapid charge separation through hole transfer to CsSnI_3 , resulting in high efficiency. Therefore, we presume that dye as well as CsSnI_3 may take part in light absorption and sensitization in a concerted way, providing a synergic effect. This issue will be the focus of our future studies.

The facile and inexpensive fabrication procedure of our cells is simple compared to other reported methods for photovoltaic devices. Powdered CsSnI_3 (about 100–300 mg) was dissolved in polar organic solvents (1.5 ml) to give transparent yellow solutions at room temperature. The solutions were injected onto the nanoporous TiO_2 electrode by a micropipette and dried under nitrogen. Complete evaporation of the solvent was confirmed with infrared spectroscopy. The unique advantage of CsSnI_3 is that a solution phase of this semiconductor material can thereby diffuse into the nanopores of TiO_2 and be stabilized as a solid phase on solvent removal. There are no by-products. Accordingly, neither high-temperature, complex chemical reactions nor high-vacuum deposition processes are required, contrary to most conventional semiconductor solar cells.

Cross-sectional scanning electron microscopy (SEM) images of our cells reveal that CsSnI_3 homogeneously permeates deep into the nanoporous TiO_2 substrate (Fig. 3a, b). The elemental mapping on the

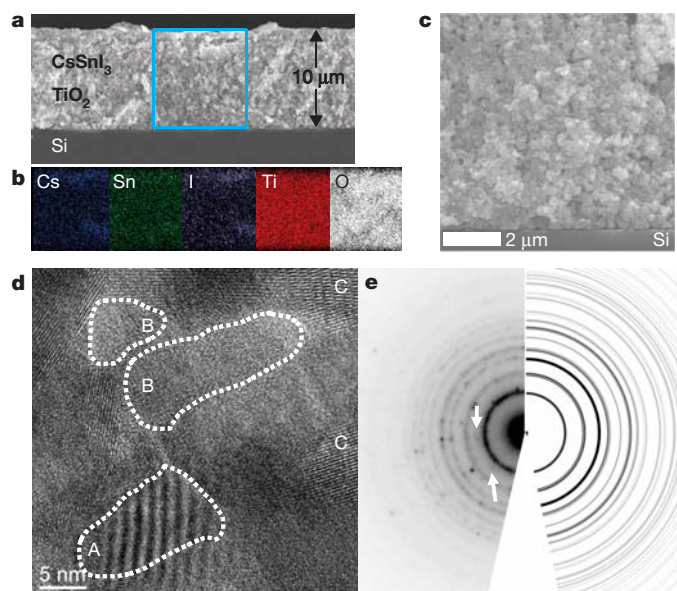


Figure 3 | Cross-sectional electron microscopy image of a $\text{CsSnI}_3/\text{TiO}_2$ cell on Si. **a**, Cross-sectional SEM image. Thickness is $\sim 10 \mu\text{m}$. **b**, Elemental mapping by energy dispersive spectroscopy for the rectangular area indicated by a blue box in **a**, showing homogeneous infiltration of CsSnI_3 into nanoporous TiO_2 . **c**, Back-scattering cross-sectional SEM image, showing no discernible contrast difference. **d**, Cross-sectional HRTEM image taken near the bottom of the cell. **A**, CsSnI_3 and TiO_2 overlapped; **B**, high (hkl) index area showing unclear lattice fringes; **C**, TiO_2 . **e**, Left panel, experimental electron diffraction pattern, indicating TiO_2 (ring patterns) and CsSnI_3 (spots indicated by white arrows between ring patterns); right panel, theoretical calculation of anatase TiO_2 .

rectangular area shown in Fig. 3a demonstrates excellent distribution of Cs, Sn and I atoms throughout the 10- μm -thick nanoporous TiO_2 (Fig. 3b). The cross-sectional back-scattering electron image shows no discernible contrast difference over the examined area, suggesting that CsSnI_3 homogeneously spreads over the entire TiO_2 area (Fig. 3c). The cross-sectional high-resolution transmission electron microscopy (HRTEM) image taken at the bottom part of the TiO_2 electrode indicates that CsSnI_3 effectively fills the TiO_2 pores, and crystallizes on the surface of the TiO_2 (Fig. 3d). The CsSnI_3 and TiO_2 phases are observed to overlap with each other, as shown in area A in Fig. 3d. Area B has too high a (*hkl*) index to give clear lattice fringes. Area C in Fig. 3d indicates TiO_2 . Figure 3e shows experimental electron diffraction patterns (left) in comparison with that of the theoretical TiO_2 (right). The ring patterns are assigned to anatase TiO_2 , and the spots (indicated by the white arrows) between the rings of TiO_2 correspond to CsSnI_3 .

To efficiently operate DSCs, the dye sensitizer that is adsorbed on the surface of the nanoporous TiO_2 transfers an electron to TiO_2 and is then regenerated by an electrolyte or p-type conductor^{1,2}. Consequently, intimate contact between the latter and dye sensitizer is essential for perfect charge separation. However, homogeneous infiltration of p-type solid HTMs throughout the n-type nanoporous materials is the main challenge for solid-state DSCs^{1,3}. Organic polymer HTMs do not efficiently infiltrate the micrometre-thick nanoporous TiO_2 , degrading performance. When monomers are used for better penetration into the pores, undesirable complex chemical reactions, such as photoelectrochemical polymerization, are required^{2,10}. The key characteristics of CsSnI_3 in our solar cells is that (1) it is solution-processable, and thus permeates throughout the entire TiO_2 structure, allowing facile charge separation and hole removal, and (2) it exhibits very large hole mobilities.

To compare the optical response of the $\text{CsSnI}_{2.95}\text{F}_{0.05}$ cell with that of a conventional Grätzel cell, the latter was prepared with N719 dye and showed $\sim 10\%$ efficiency¹⁸. The optical absorption spectrum of the $\text{CsSnI}_{2.95}\text{F}_{0.05}$ -containing cell, obtained in transmission mode, reveals a well-defined edge at 789 nm, which is significantly red-shifted from that of the Grätzel cell¹⁸ with N719 dye, at 667 nm (Fig. 4a). This observation indicates that our cell absorbs red and near-infrared light more efficiently than the Grätzel cell. Note that lack of sunlight absorption in the red and near-infrared regions has been a challenge for typical ruthenium-based dyes. Figure 4b compares the incident photon-to-current conversion efficiency (IPCE) as a function of excitation wavelength for the $\text{CsSnI}_{2.95}\text{F}_{0.05}$ cell with the Grätzel cell. The IPCE spectrum is a measure of the light response of photovoltaic

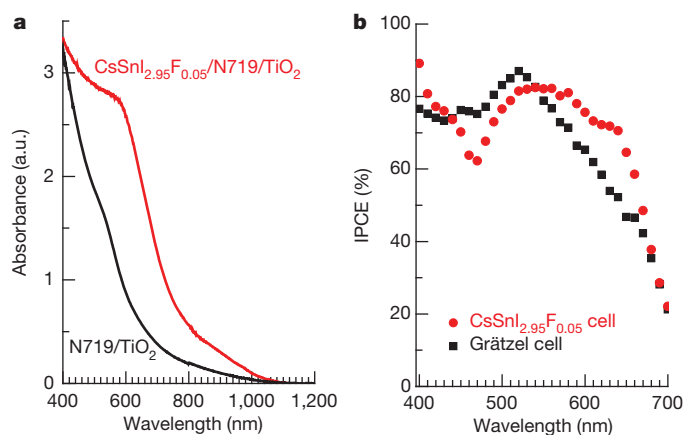


Figure 4 | Optical response of the $\text{CsSnI}_{2.95}\text{F}_{0.05}$ cell and a conventional Grätzel cell. **a**, Optical absorbance spectra of the devices consisting of $\text{CsSnI}_{2.95}\text{F}_{0.05}/\text{N719 dye}/\text{TiO}_2$ (red line) and N719 dye/ TiO_2 (liquid electrolyte was not added here) (black line). **b**, The IPCE spectrum as a function of the wavelength of monochromatic light that impinges on the $\text{CsSnI}_{2.95}\text{F}_{0.05}$ cell (filled circles) in comparison with that of the N719-dye-containing Grätzel cell (filled squares).

devices, which is directly related to the short-circuit current. In the 550–670 nm spectral range, our cell produces a higher and broader photocurrent density in the external circuit under monochromatic illumination (per photon flux). Note that the upper limit of our IPCE measurement setting is 700 nm, resulting in a sharp drop beyond 670 nm.

The pristine CsSnI_3 cell gave very good photocurrent density–voltage (*J*–*V*) characteristics—open-circuit voltage (V_{OC}), fill factor (FF), short circuit current density (J_{SC}) and overall power conversion efficiency (η)—as a solid-state solar cell: $V_{\text{OC}} = 0.638\text{ V}$, $\text{FF} = 66.1\%$, $J_{\text{SC}} = 8.82\text{ mA cm}^{-2}$, $\eta = 3.72\%$ ($\eta = V_{\text{OC}} \times J_{\text{SC}} \times \text{FF}$) (Fig. 5). To improve *J*–*V* characteristics, we studied the effect of fluorine doping on CsSnI_3 . The optimum molar concentration of 5% fluorine doping gave a remarkable increase in J_{SC} , which reached 12.2 mA cm^{-2} , resulting in a 1.5-fold larger value of η , 5.62%. Further improvement was obtained by introducing SnF_2 into $\text{CsSnI}_{2.95}\text{F}_{0.05}$. The SnF_2 doping process was simple. The desired amounts of SnF_2 and $\text{CsSnI}_{2.95}\text{F}_{0.05}$ powders were added to polar organic solvents with stirring. The resulting solutions were injected onto the nanoporous TiO_2 electrodes. The $\text{CsSnI}_{2.95}\text{F}_{0.05}$ sample doped with 2% SnF_2 provided a 29% and 21% increase in J_{SC} (15.7 mA cm^{-2}) and η (6.81%), respectively, compared to the $\text{CsSnI}_{2.95}\text{F}_{0.05}$ sample. The optimum molar concentration of SnF_2 doping in $\text{CsSnI}_{2.95}\text{F}_{0.05}$ was found to be 5%. For the corresponding cell, the TiO_2 nanoporous film was pre-treated by a fluorine plasma etching process to increase the size of the nanopores and nanochannels as described in the literature¹⁸. This also possibly helps to reduce surface states and charged particle recombination¹⁸. The resulting cell showed very good *J*–*V* characteristics: $J_{\text{SC}} = 17.4\text{ mA cm}^{-2}$, $V_{\text{OC}} = 0.730\text{ V}$, $\text{FF} = 72.9\%$, $\eta = 9.28\%$.

To fully employ the photon flux absorbed, we applied two layers of the three-dimensional inverse photonic crystal ZnO (ref. 19) over the counter electrode of the same cell. Each layer of the photonic crystal had a different hole diameter—values of 375 nm and 410 nm were used. The corresponding cell exhibited $J_{\text{SC}} = 19.2\text{ mA cm}^{-2}$, $V_{\text{OC}} = 0.732\text{ V}$, $\text{FF} = 72.7\%$ and $\eta = 10.2\%$. When a mask was applied on the cell, $\eta = 8.51\%$ was observed (Supplementary Information). The observed efficiency is the highest among any kind of dye-sensitized solar cell free of liquid electrolyte, and is close to that of

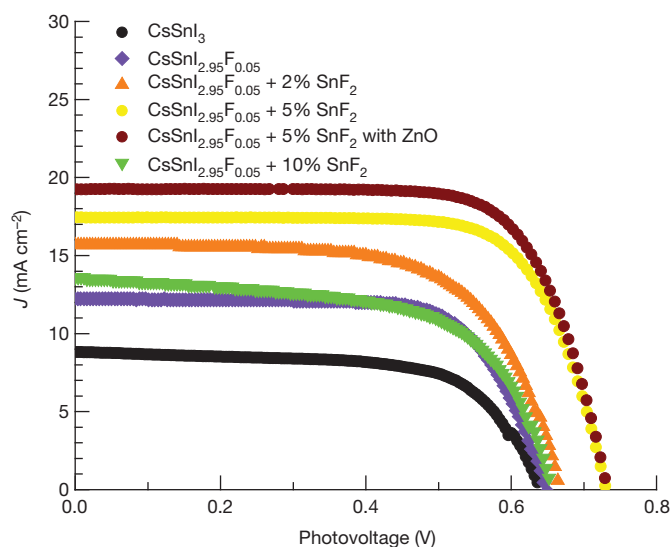


Figure 5 | Photocurrent density–voltage (*J*–*V*) characteristics of the solar cell devices under irradiation of 100 mW cm^{-2} simulated AM 1.5 sunlight. These solar cells incorporated CsSnI_3 and $\text{CsSnI}_{2.95}\text{F}_{0.05}$ doped with SnF_2 . Fluorine and SnF_2 doping increase photocurrent density, resulting in an improved power conversion efficiency (η). Application of three-dimensional ZnO photonic crystal layers further enhances the photocurrent density, and gives the highest value of η (10.2%) for the cell of $\text{CsSnI}_{2.95}\text{F}_{0.05}$ doped with SnF_2 .

the high performance Grätzel cell ($\eta \approx 11\%$). The samples with 10% SnF_2 resulted in a reduction in J_{SC} (measured value, 13.6 mA cm^{-2}) and η (measured value, 5.46%). Compounds of the type $\text{CH}_3\text{NH}_3\text{PbX}_3$ ($\text{X} = \text{Br}, \text{I}$), which are isostructural with CsSnI_3 , have been used as visible light sensitizers adsorbed on the TiO_2 surface for photovoltaic cells. These cells, however, employed organic liquid electrolyte containing LiX/X_2 redox couples, and exhibited low efficiency²⁰.

We have demonstrated the superior performance of the p-type inorganic high-hole-mobility semiconductor $\text{CsSnI}_{3-x}\text{F}_x$ and its ability to replace the problematic organic liquid electrolytes in dye-sensitized solar cells (DSCs). We refer to these solid-state solar cells as solid-state DSCs (SSDSCs). Unlike liquid electrolytes with their complex chemistry, crystalline inorganic solids promise long-term stability, and also enable theoretical calculations based on electronic structure and further improvement of their photovoltaic properties. The new solar cells described here are all-solid-state, inorganic systems. The $\text{CsSnI}_{3-x}\text{F}_x$ compounds consist of inexpensive, abundant elements and are solution-processable at room temperature, allowing for low-cost processing. Our new cell reaches a conversion efficiency of $\sim 10.2\%$ (8.51% with a mask), and is the first example of an all solid-state dye-sensitized solar cell system that may eventually exceed the performance of a liquid electrolyte Grätzel cell. With further optimization and new dyes¹¹, much higher efficiencies are likely. This work opens up the possibility of semiconducting solid materials becoming 'state of the art' and promoting much higher efficiencies than have been possible with conventional DSCs.

METHODS SUMMARY

Synthesis of $\text{CsSnI}_{3-x}\text{F}_x$. Pure $\text{CsSnI}_{3-x}\text{F}_x$ ($0 \leq x \leq 1$) compounds were achieved by heating a stoichiometric mixture of CsI , SnI_2 and SnF_2 in an evacuated Pyrex or fused silica tube at 450°C for 30 min, followed by quenching to room temperature. The ground powders ($\sim 100\text{--}350 \text{ mg}$) were dissolved/dispersed in anhydrous polar organic solvents (1.5 ml): *N,N*-dimethylformamide, acetonitrile and methoxyacetonitrile. For SnF_2 doping, appropriate ratios of $\text{CsSnI}_{2.95}\text{F}_{0.05}$ and SnF_2 powders were stirred in the same organic solvents.

TiO_2 electrode preparation and device assembly. TiO_2 electrode preparation, a fluorine plasma etching process, and device assembly are described in the literature¹⁸. The solutions of $\text{CsSnI}_{3-x}\text{F}_x$ (with SnF_2 if necessary) were injected into the cell by a micropipette and dried. The ZnO photonic crystals were prepared as described in the literature¹⁹. They were attached on the top of the counter-electrode if necessary.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 February; accepted 8 March 2012.

- Hagfeldt, A., Boschloo, G., Sun, L. C., Kloo, L. & Pettersson, H. Dye-sensitized solar cells. *Chem. Rev.* **110**, 6595–6663 (2010).
- Grätzel, M. Recent advances in sensitized mesoscopic solar cells. *Acc. Chem. Res.* **42**, 1788–1798 (2009).

- Yanagida, S., Yu, Y. H. & Manseki, K. Iodine/iodide-free dye-sensitized solar cells. *Acc. Chem. Res.* **42**, 1827–1838 (2009).
- Koh, J. K., Kim, J., Kim, B., Kim, J. H. & Kim, E. Highly efficient, iodine-free dye-sensitized solar cells with solid-state synthesis of conducting polymers. *Adv. Mater.* **23**, 1641–1646 (2011).
- Daenke, T. *et al.* High-efficiency dye-sensitized solar cells with ferrocene-based electrolytes. *Nature Chem.* **3**, 211–215 (2011).
- Wang, M. K. *et al.* An organic redox electrolyte to rival triiodide/iodide in dye-sensitized solar cells. *Nature Chem.* **2**, 385–389 (2010).
- Liu, X. Z. *et al.* An efficient organic-dye-sensitized solar cell with in situ polymerized poly(3,4-ethylenedioxythiophene) as a hole-transporting material. *Adv. Mater.* **22**, E150–E155 (2010).
- Jiang, K. J. *et al.* Photovoltaics based on hybridization of effective dye-sensitized titanium oxide and hole-conductive polymer P3HT. *Adv. Funct. Mater.* **19**, 2481–2485 (2009).
- Bach, U. *et al.* Solid-state dye-sensitized mesoporous TiO_2 solar cells with high photon-to-electron conversion efficiencies. *Nature* **395**, 583–585 (1998).
- Bisquert, J. Dilemmas of dye-sensitized solar cells. *ChemPhysChem* **12**, 1633–1636 (2011).
- Yella, A. *et al.* Porphyrin-sensitized solar cells with cobalt (II/III)-based redox electrolyte exceed 12 percent efficiency. *Science* **334**, 629–634 (2011).
- Cai, N. *et al.* An organic D- π -A dye for record efficiency solid-state sensitized heterojunction solar cells. *Nano Lett.* **11**, 1452–1456 (2011).
- O'Regan, B., Lenzmann, F., Muis, R. & Wience, J. A solid-state dye-sensitized solar cell fabricated with pressure-treated P25-TiO_2 and CuSCN : analysis of pore filling and IV characteristics. *Chem. Mater.* **14**, 5023–5029 (2002).
- Bandara, J. & Weerasinghe, H. Solid-state dye-sensitized solar cell with p-type NiO as a hole collector. *Sol. Energy Mater. Sol. Cells* **85**, 385–390 (2005).
- Kroon, J. M. *et al.* Nanocrystalline dye-sensitized solar cells having maximum performance. *Prog. Photovolt. Res. Appl.* **15**, 1–18 (2007).
- Shum, K. *et al.* Synthesis and characterization of CsSnI_3 thin films. *Appl. Phys. Lett.* **96**, 221903 (2010).
- Grätzel, M. Photoelectrochemical cells. *Nature* **414**, 338–344 (2001).
- Lee, B., Buchholz, D. B., Guo, P. J., Hwang, D. K. & Chang, R. P. H. Optimizing the performance of a plastic dye-sensitized solar cell. *J. Phys. Chem. C* **115**, 9787–9796 (2011).
- Lee, B. *et al.* Materials, interfaces, and photon confinement in dye-sensitized solar cells. *J. Phys. Chem. B* **114**, 14582–14591 (2010).
- Kojima, A., Teshima, K., Shirai, Y. & Miyasaka, T. Organometal halide perovskites as visible-light sensitizers for photovoltaic cells. *J. Am. Chem. Soc.* **131**, 6050–6051 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors acknowledge support for this collaborative research: NSF-DMR 0843962 for R.P.H.C.; DOE Energy Frontier Research Center, ANSER, DE-SC0001059 for B.H.L., J.H. and M.G.K.; the Initiative for Energy and Sustainability at Northwestern (ISEN) for I.C. Device testing and measurements were done in the ANSER Facilities and materials characterization was performed in the NSFMRSEC Facilities (DMR-1121262).

Author Contributions I.C. and M.G.K. conceived and designed the experiments and prepared the manuscript. I.C. synthesized materials. R.P.H.C. and B.L. designed and fabricated the solar cells, I.C. and B.L. performed measurements. J.H. collected TEM data. I.C., B.L., R.P.H.C. and M.G.K. discussed the results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.G.K. (m-kanatzidis@northwestern.edu).

METHODS

TiO₂ electrode preparation. High-purity anatase TiO₂ nanoparticles and their paste form were prepared as described in the literature¹⁸. The obtained paste was spread on the fluorine-doped SnO₂ glass substrate (FTO, Pilkington, TEC 8 glass, 8 Ω per square, 2.3 mm thick) using a doctor blade, followed by gradual calcination to remove residual polymers under an air flow at 150 °C for 15 min, at 320 °C for 10 min and at 500 °C for 30 min, giving a pure anatase TiO₂ nanoparticle film with a flat and smooth surface. A fluorine plasma etching process was applied as described in the literature if necessary¹⁸. The film thickness was determined by a surface profiler (Tencor P-10). The ZnO photonic crystals were prepared as described in the literature¹⁹. They were attached on the top of the counter-electrode if necessary.

Device assembly. The TiO₂ electrode was immersed in an ethanol solution containing purified 3×10^{-4} M (*cis*-diisothiocyanato-bis(2,2'-bipyridyl-4,4'-dicarboxylato) ruthenium(II) bis(tetrabutylammonium)) (N719, Solaronix) for 18 h at room temperature, followed by rinsing with ethanol and drying under a nitrogen flow. The counter-electrode was prepared by coating FTO glass with a thin layer of a 5 mM H₂PtCl₆ solution (in isopropanol), followed by heating at 400 °C for 20 min. The two electrodes were sealed together with thermal melt polymer film (Surlyn, DuPont). The typical active area of the cell was ~ 0.2 – 0.3 cm². The exact area of photoanode was calibrated by an optical scanner under a resolution of 600 dots per inch. The CsSnI_{3-x}F_x without SnF₂ solutions were injected into the cell by a micropipette and dried.

Device characterization. The devices were evaluated under 100 mW cm⁻² standard air mass 1.5 global (AM 1.5G) sunlight simulation with a class A solar cell analyser (Spectra Nova Tech.). A silicon solar cell fitted with a KG3 filter tested and certified by the National Renewable Energy Laboratory (NREL) was used for calibration. The KG3 filter accounts for the different light absorptions of the dye-sensitized solar cell and silicon, and ensures that the spectral mismatch correction factor approaches unity. Incident photon-to-current conversion efficiency was measured using a monochromator (Jobin-Yvon fluorescence spectrometer) and a potentiostat

(CH Instruments 1202). Electrochemical impedance spectroscopy was performed under the same light illumination with an impedance analyser (Solartron 1260) and a potentiostat (Solartron 1287) when the device was tested at its V_{OC} . An additional low amplitude modulation sinusoidal voltage of 10 mV_{rms} was also applied between the anode and cathode of the device over the frequency range of 0.05–150 kHz.

Electron microscopy. SEM images were taken with a Hitachi S-4800II SEM. TEM investigations were carried out in a JEOL 2100F microscope operating at 200 keV accelerating voltage.

CsSnI₃ characterization. Hall effect measurement. The CsSnI₃ sample was subjected to a flow of nitrogen gas to protect it from oxidation during measurements. A d.c. current source (6220 Keithley Instruments) was used to supply a constant 100 mA current through the sample while an external magnetic field up to ± 1.4 T was applied in steps of 0.25 T. The voltage response, both in the positive and negative directions of the magnetic field, was recorded at each respective step by a nanovoltmeter (2182A Keithley Instruments) and the half-difference of the two signals was recorded as the Hall voltage.

Charge-transport properties. The ingot was cut and polished into a rectangular shape with dimensions of ~ 2 mm \times 3 mm \times 8 mm under N₂ atmosphere. Electrical conductivity and Seebeck coefficient were measured simultaneously under He atmosphere from room temperature to 650 K on a ULVAC-RIKO ZEM-3 instrument system. The mechanical electrodes made of Rh/Pt and Pt were directly contacted to the ingot with no use of a chemical paste.

Optical absorption spectroscopy. Optical diffuse reflectance measurements were performed at room temperature using a Shimadzu UV-3101 PC spectrometer operating in the 200–2,500 nm region. The reflectance versus wavelength data generated were used to estimate the bandgap of the material by converting reflectance to absorption data according to the Kubelka-Munk equation: $\alpha/S = (1 - R)^2 (2R)^{-1}$, where R is the reflectance and α and S are the absorption and scattering coefficients, respectively.

Brain-wide neuronal dynamics during motor adaptation in zebrafish

Misha B. Ahrens^{1,2}, Jennifer M. Li¹, Michael B. Orger³, Drew N. Robson¹, Alexander F. Schier¹, Florian Engert¹ & Ruben Portugues¹

A fundamental question in neuroscience is how entire neural circuits generate behaviour and adapt it to changes in sensory feedback. Here we use two-photon calcium imaging to record the activity of large populations of neurons at the cellular level, throughout the brain of larval zebrafish expressing a genetically encoded calcium sensor, while the paralysed animals interact fictively with a virtual environment and rapidly adapt their motor output to changes in visual feedback. We decompose the network dynamics involved in adaptive locomotion into four types of neuronal response properties, and provide anatomical maps of the corresponding sites. A subset of these signals occurred during behavioural adjustments and are candidates for the functional elements that drive motor learning. Lesions to the inferior olive indicate a specific functional role for olivocerebellar circuitry in adaptive locomotion. This study enables the analysis of brain-wide dynamics at single-cell resolution during behaviour.

The generation of motor output and the influence of sensory input on future motor programs engage neural activity in many neurons across multiple brain regions. However, past measurements of neural activity during behaviour have been hampered by the inability to monitor exhaustively all neurons in the brain of a behaving animal. Although it is possible to record activity from behaving animals^{1–6}, the large size and opacity of the vertebrate brain constrains experimenters to focus on small fractions of the total number of neurons. Here we develop a preparation in which neuronal activity can be monitored anywhere in the brain using two-photon calcium imaging in paralysed larval zebrafish that interact with a virtual environment and adjust their behaviour to changes in visual feedback.

When visual feedback following a motor command does not meet expectation, animals can learn to adapt the strength of subsequent motor commands. In the past this has been studied in controlled laboratory settings by perturbing visual feedback in the context of insect flight^{7–9}, the vestibulo-ocular reflex^{10,11} and reaching movements^{12,13}. Here we study adaptive control of locomotion in larval zebrafish¹⁴. This animal swims in discrete swim bouts during which the visual environment moves relative to its retina. One hypothesis is that this optic flow is used as a measure of displacement and serves to tune the strength of future motor commands to the desired travel distance^{7–9}. Such sensorimotor recalibration is particularly important during the optomotor response^{15,16}, in which animals move in the direction of motion of the visual surround—thereby stabilizing their location in the presence of, for example, water flow—a response that occurs in many animal species. If motor output is not correctly calibrated to visual feedback, a fish may systematically overshoot or undershoot the desired travel distance, instead of stabilizing its location. Sensorimotor recalibration is necessary for accurate locomotion because the rate of optic flow following a motor command is affected by temperature-dependent changes in muscle strength, the viscosity of the water and the distance of objects from the retina.

To examine neural dynamics across brain areas that drive sensorimotor recalibration, we developed a system to study neural activity at cellular resolution^{17,18}, using two-photon microscopy¹⁹, anywhere in

the brain²⁰ during closed-loop optomotor behaviour in larval zebrafish. These animals have a small and transparent brain that is readily accessible for optogenetic recording and stimulation^{21,22}, electrophysiology²³ and single-cell ablation²⁴. To remove motion artefacts^{25,26}, we developed a swim simulator for completely paralysed larvae (Fig. 1a). Motor commands, or ‘fictive swims’, are recorded at the motor neuron level^{8,27,28} (Fig. 1c, d) and translated, in real time, into visual feedback that resembles the optic flow of freely swimming fish (Methods). This constitutes a fictively driven virtual-reality setup. Simultaneously, a two-photon microscope scanning over a transgenic fish expressing GCaMP2 (ref. 29) in almost all neurons^{20,30} allows activity to be monitored throughout the brain at single-neuron resolution. As the experimenter is in complete control of the visual feedback, this allowed us to study neural dynamics during visually guided motor adaptation throughout the brain at the cellular level.

Fictive motor adaptation

To study motor adaptation, we used a closed-loop paradigm and simulated a one-dimensional environment in which the fish is swept backwards by a virtual water flow, a motion that the fish was able to compensate for by swimming forwards, as in the optomotor response. In the fictive virtual-reality setup, this corresponds to a whole-field visual stimulus that is moving forwards but that can be momentarily accelerated backwards by a fictive swim of the fish (Fig. 1b; Methods), so that the fish can stabilize its virtual location over time. Remarkably, paralysed larval zebrafish behaved readily in this closed-loop paradigm, showing similar behaviour to freely swimming fish that are exposed to whole-field motion²⁴, and were not noticeably compromised by the absence of vestibular, proprioceptive and somatosensory feedback that accompanies unrestrained swimming.

An important free parameter in this closed-loop paradigm is the feedback gain^{10,11,12,14,31}—the factor that translates the strength of the fictive swim signal to the change in stimulus velocity (Methods). The higher the feedback gain, the greater the velocity change following a motor command (for example, dashed red line in Fig. 1b), so that high feedback gain corresponds to a ‘strong virtual fish’ and low feedback

¹Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA. ²Computational and Biological Learning Laboratory, Department of Engineering, Cambridge University, Trumpington Street, Cambridge CB2 1PZ, UK. ³Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Avenida Brasília, Doca de Pedrouços, 1400-038 Lisboa, Portugal.

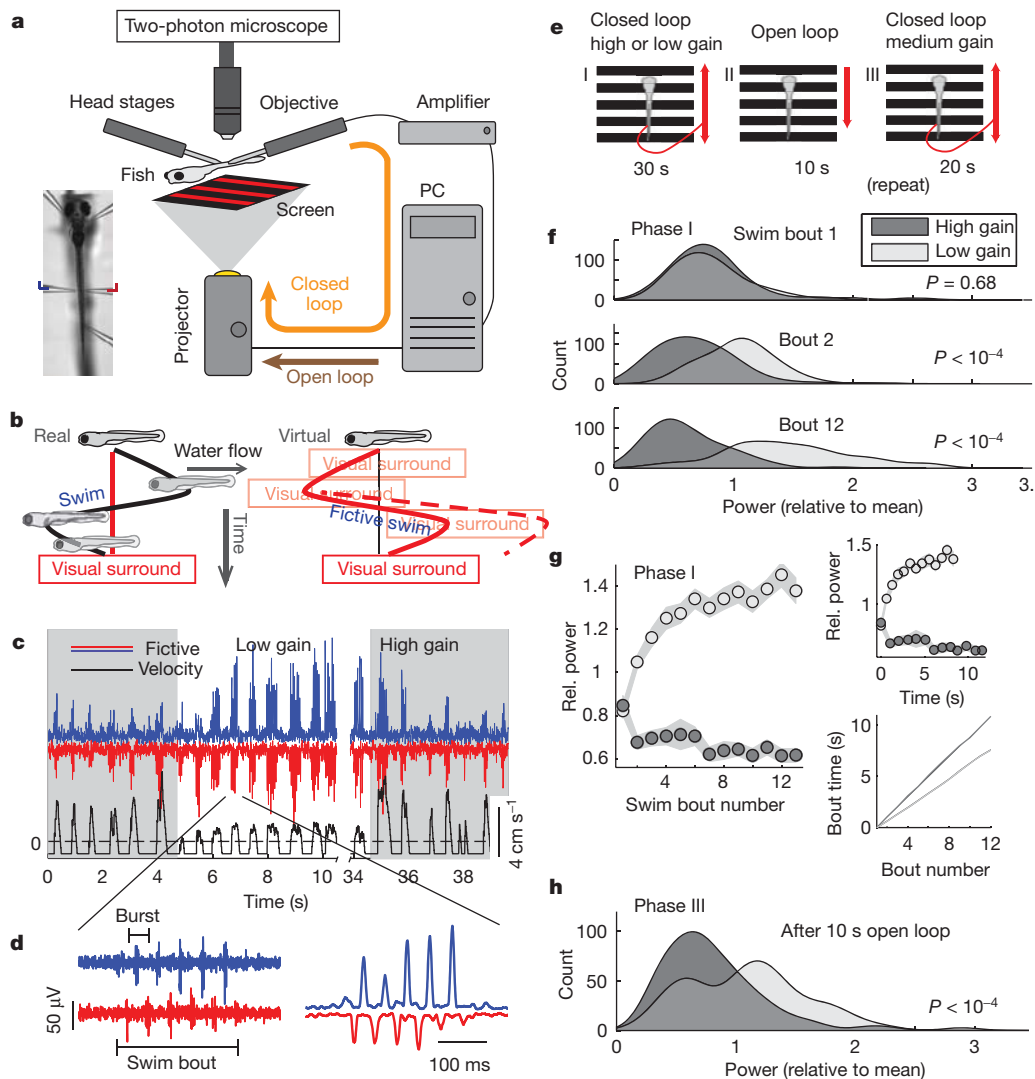


Figure 1 | Experimental setup and fictive motor adaptation. **a**, Schematic of the setup. Photograph of a paralysed larval zebrafish (left) in the experimental setup (right), supported by pipettes, two of which are recording pipettes. **b**, Illustration of the virtual motor adaptation assay. Left, trajectory of a fish executing one swim bout against a water current (black trajectory) in the presence of a visual surround (red). Right, simulation of this behaviour in the virtual environment, in which the visual surround is moved and the fish is stationary. The visual surround is accelerated backwards when a fictive swim occurs. The trajectory that would occur if the feedback gain were higher is shown by a dashed red line. **c**, Fictive motor adaptation. Fictive swim vigour (blue and red traces for left and right channels, respectively) and stimulus velocity (black trace) plotted over time. High and low feedback gain epochs are

gain to a 'weak virtual fish'. Accurate motor control would require the motor output to adapt to the feedback gain. Periodically switching the feedback gain between low and high values in the virtual environment resulted in compensatory changes in motor output: a change to a lower gain resulted in the gradual increase of the amplitude and duration of the fictive swim signals (Fig. 1c and Supplementary Movie 1; a weak fish sends more impulses to the muscles), whereas a switch to a higher gain setting led to an incremental decrease (a strong fish sends fewer impulses to the muscles). This behaviour was tested in more detail in the scheme shown in Fig. 1e, which was repeated up to 50 times per fish. We analysed the power of motor nerve bursts, equivalent to the number of fictive tail oscillations (Fig. 1d), in each of the first 12 swim bouts that occurred during 30 s of motor adaptation to either high or low feedback gain (Fig. 1e, phase I). The first swim bouts following a switch from low to high gain, or from high to low gain,

were indistinguishable (t -test, $P > 0.5$; Fig. 1f). This implies that fish do not adjust their motor output once a motor command has been issued, despite the presence of immediate visual feedback. Starting at the second swim bout, the power diverges in the high and low gain conditions ($P < 10^{-4}$). Behavioural adjustment plateaus after about ten bouts, which corresponds to approximately 7–10 s (Fig. 1g; Supplementary Fig. 4).

To determine whether the larvae are learning a new sensorimotor transformation or merely responding to different patterns of visual stimulation during the high- and low-gain periods, fish were exposed to a 10-s 'rest' period (Fig. 1e, phase II) during which constant-velocity backward gratings were shown in open loop, a stimulus that tends to inhibit swimming, followed by a closed-loop 'test' period of medium feedback gain (Fig. 1e, phase III). We found that the strength of the first swim bouts in the 'test' period was determined by the gain

setting during the preceding adaptation session (phase I), which shows that the retention of the increased or decreased locomotor drive outlasts 10 s of fixed optic flow (Fig. 1h and Supplementary Fig. 5). Thus, motor adaptation in larval zebrafish is not merely a response to different patterns of visual stimulation, but instead involves a short-term learned change in the sensorimotor transformation.

Brain-wide functional imaging during behaviour

After verifying that neural activity in the reticulospinal system is modulated by locomotor drive (Supplementary Fig. 6a–f), as suggested by previous studies²⁴, we next looked for signals relating to adaptive motor control throughout the entire brain (Supplementary Movie 2). We generated a transgenic fish expressing the genetically encoded calcium indicator GCaMP2 (ref. 29) driven by the panneuronal *elavl3* (previously known as HuC) promoter^{20,30} (Fig. 2a). We used a paradigm in which 30 s of high gain alternated with 30 s of low gain, without open-loop intermissions. Behavioural variables such as swim frequency, number of bursts and power changed in an analogous manner (Supplementary Fig. 4c). A single *z* plane was imaged for six repetitions of gain switches. The brain volume that can be covered in a single fish depends on the duration of the paradigm and the size of the imaging plane. With relatively short assays (about 2 min) the entire brain of a single fish can be imaged in one experiment. We chose instead to use a longer assay—10 min—to cope with our relatively complicated behavioural paradigm and the low signal-to-noise ratio of GCaMP2. Thus, we sampled on average 20% of each fish's brain and created a composite brain for final analysis.

Data analysis was automated and carried out as follows. Every experiment generated a number of fluorescence movies with associated fictive swim recordings and information about the stimulus. A custom-written signal-identification and signal-localization algorithm extracted fluorescence time series from single neurons or $4 \times 4\text{-}\mu\text{m}$ regions of neuropil (Fig. 2c–e; Methods). These fluorescence time series (Fig. 2e) were then related to the stimulus and behavioural traces (Fig. 2f) using the methods described below. Finally, to identify the regions of interest (ROIs) of multiple fish with anatomical loci in a reference brain, all imaged planes were mapped using an image-registration algorithm (Supplementary Movie 3) to a high-resolution reference brain of a 6-days-post-fertilization (d.p.f.) larva. Although small variations existed between brains, the large number of landmarks made reasonable localization between different fish possible (within about $25\text{ }\mu\text{m}$; Methods).

Measurement of motor-related activity

To search the brain for neural activity related to motor output, we first needed to solve the problem that in closed loop, it is not possible to easily distinguish motor- from visual-related activity, as both are directly linked in this setting. A period of open-loop stimulus presentation was added to the paradigm (Fig. 2e, f; yellow area), during which the stimulus experienced by the animal during a preceding closed-loop period (Fig. 2e; black bar) was repeated. Activity of visually driven neurons during this 'replay' period will resemble the activity of the preceding period, and this is formalized by using the correlation coefficient of fluorescence (CCFF) during and before replay, as a measure of the degree of visually driven activity. On the other hand, activity of motor-related neurons will instead correlate to motor output, and this is formalized by CCFFM, the correlation coefficient between motor output and fluorescence during replay, as a measure of motor-related activity (Methods). High $|\text{CCFFM}|$ indicates activity related to locomotion (for example, Fig. 2e), and high CCFF indicates visually driven activity.

As in Fig. 2f, swimming behaviour of most fish became erratic during open-loop stimulus replay. This observation shows that the behavioural state of an animal strongly depends on the presence of appropriate sensory feedback following a motor command.

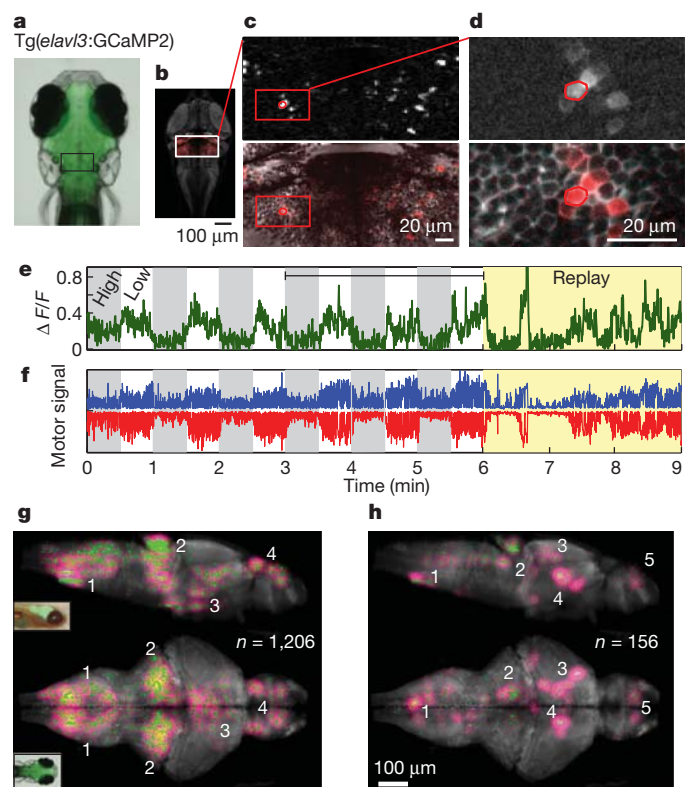


Figure 2 | Functional imaging during adaptive motor control in larval zebrafish. **a**, Micrograph of a transgenic larval zebrafish with pan-neuronal GCaMP2 expression under control of the *elavl3* promoter. The area scanned in **c–e** is indicated by a box. **b**, Automatic localization of imaged plane on a reference brain (Methods). **c**, Detected signal (top; Methods) overlaid on anatomical image (bottom). The neuron of interest selected in **d** and **e** is indicated by a red circle. **d**, Correlation map (top, correlation coefficient of ROI signal with all pixels in the red rectangle; Methods) overlaid on anatomy (bottom). The example neuron has been hand segmented. **e**, Fluorescence time series of example neuron. Grey, high feedback gain; white, low feedback gain. Yellow, open-loop stimulus playback from the previous 3 minutes (black bar). **f**, Fictive locomotor drive is boosted during low gain periods. The fluorescence time series in **e** is strongly correlated with the fictive swim signal (yellow replay period: $|\text{CCFFM}| = 0.58 > \text{CCFF} = 0.21$). **g**, Areas in the brain with activity strongly correlating to fictive locomotion ($|\text{CCFFM}| > 0.5$; n is the number of sites satisfying the criterion). Green dots, location of identified sites; magenta–yellow gradient, spatial uncertainty (caused by mapping 32 brains to one reference brain), scaled by sampling density. Units were found in 1, areas of the hindbrain, including the inferior olive; 2, in the cerebellum and anterior hindbrain; 3, in the NMLF and pretectum area; 4, in the forebrain. Top panel, side view; bottom panel, top view. **h**, Areas correlating with visual stimulation but not motor output ($\text{CCFF} > |\text{CCFFM}|$, $\text{CCFF} > 0.2$; Methods). 1, hindbrain; 2, cerebellum; 3, tectum; 4, pretectum; 5, forebrain. Top panel, side view; bottom panel, top view.

Figure 2g shows the density of neurons in the reference brain ($n = 32$ fish) whose activity was strongly correlated to fictive motor output during replay (high $|\text{CCFFM}|$; Methods). Clusters of such neurons can be seen in the caudal hindbrain, including the inferior olive, in the cerebellum, near the nucleus of the medial longitudinal fasciculus (NMLF) and pretectum, and in the forebrain. Asymmetries in the anatomical maps may arise from the limited sensitivity of GCaMP2 and the limited sampling of the brain. Figure 2h shows the density of neurons whose activity correlates with visual input (high CCFF; Methods). Here we used a more liberal criterion because only a small fraction of all possible visual input could be sampled, so that this map is by necessity less complete than that in Fig. 2g. Nevertheless, densities were found in the area of the pretectum, the tectum, the cerebellum and the hindbrain (including the inferior olive). Thus, regions throughout the brain that are involved

in locomotion could be identified through correlational analysis of neural activity during behaviour.

Phase-space representation of network activity

As a first step towards understanding the dynamics that occur during motor adaptation, activity of all identified sites across all fish was visualized by embedding it in a three-dimensional phase space using principal components analysis (PCA)^{32,33} (traces averaged over six low-high gain repetitions; see Methods).

As expected, the trajectory loops back to the starting point (Fig. 3a, c), reflecting the periodicity of the neural activity induced by the paradigm, which consisted of repeating periods of high and low feedback gain. The velocity through the principal component space is initially high after a change in feedback gain, then slows down (Fig. 3d) to reach one of two approximately-steady states (β and δ in Fig. 3c). Notably, the periods of fast change in network space (α and γ) coincide with the period of behavioural change (Fig. 1g). Indeed, the first two temporal principal components shown in Fig. 3b reflect steady-state activity (TCP1) and transient activity (TCP2) after a decrease in feedback gain. In summary, network activity evolves quickly following a change in feedback gain, and then settles into one of two steady states depending on the setting of the feedback gain. Network changes coincide with changes in behaviour, and steady network states correspond to periods of stable behaviour. To determine what neural activity induces the two transient and the two steady phases, we next looked for neurons that showed correlated activity with these four phases, that is, during the four phases α – δ in Fig. 3d.

Neural correlates of adaptive motor control

Motor-related neurons (phase β)

Neurons exhibiting raised activity during the low-gain, high-locomotor-drive phase were termed ‘motor’ related neurons (average fluorescence $F_\beta > F_\delta$, paired *t*-test on six repetitions, $P < 0.005$). The neurons of Figs 2e and 4a are two examples. Activity of these two neurons was more related to locomotion than to visual input, as determined by $|CCFM| > CCFF$ during replay (Fig. 2e: $|CCFM| = 0.58| > CCFF = 0.21$, Fig. 4a: $|CCFM| = 0.154| > CCFF = 0.052$). This was the case for almost all members of the ‘motor’ population (Fig. 4g). The population average of activity of ‘motor’ neurons is shown in Fig. 4f.

‘Motor’ units were found in areas shown in Fig. 5a: in the posterior hindbrain^{34,35}, with a particularly dense concentration just caudal to the cerebellum, and in the inferior olive; in the reticulospinal system (Supplementary Fig. 6), as suggested by previous studies^{24,25}; throughout the cerebellum, particularly in the corpus, including in areas of the Purkinje (larger, more dorsal cell bodies) and in granule cell (smaller, deeper cell bodies) layers^{36–38}, and in the deep cerebellum (more lateral); in the midbrain, ventral to the optic tectum, near the NMLF and in the pretectum^{38,39}; in the habenula; and in the pallidum.

Gain-decrease-related neurons (phase α)

Neuronal populations that are active after a decrease in gain may be responsible for driving the animal into a state of high locomotor drive. Signals during this period are also consistent with the detection of discrepancies between expected and received visual feedback, which are thought to drive many forms of adaptive behaviour^{10,11,12,31,40,41}. An example from the ‘gain-down’ population ($F_\alpha > F_\beta$, $F_\alpha > F_\delta$; two *t*-tests on six gain repetitions, $P < 0.005$), at the ventrolateral border of the cerebellum, is shown in Fig. 4b: after a gain decrease, this neuron shows transient activity that returns to baseline while the increase in locomotor drive is maintained. The stimulus replay period shows that the neuron is not visually driven, but is instead functionally related to motor output ($|CCFM| = 0.32| > CCFF = 0.15$), as are most other neurons of this population (Fig. 4g).

Gain-down units were found in the cerebellum and in the inferior olive (Fig. 5b). In the cerebellum they appeared in both the areas of the Purkinje and granule cell layers, and in the deep cerebellum. In the dorsal areas they appeared medial, and in the deep cerebellum they appeared lateral. The anatomical localization of gain-down units to the cerebellum is consistent with findings in mammals, in which the cerebellum is a locus of motor learning^{10,11}. Some units were also found in the hindbrain just caudal to the cerebellum.

Gain-increase-related neurons (phase γ)

Figure 4c shows calcium signals of a ‘gain-up’ neuron in the optic tectum that shows transient signals after an increase in gain ($F_\gamma > F_\beta$, $F_\gamma > F_\delta$). This neuron is mainly driven by visual input, as the calcium trace during the replay period resembles the trace during the matched preceding period ($|CCFM| = -0.21| < CCFF = 0.59$, representative of the population, see Fig. 4g). Not many neurons with the gain-up property were found ($n = 39$), but a concentration was found in the inferior olive (Fig. 5b), and most of these neurons were visually driven (Fig. 4g).

Motor-off-related neurons (phase δ)

A fourth class of neurons showed raised activity during periods of weak locomotion or absence of locomotion, termed the ‘motor-off’ class ($F_\delta > F_\beta$). Figure 4d shows an example of such a neuron, in the dorsal hindbrain, whose activity is elevated during periods of high gain and low locomotor drive. Notably, during the stimulus replay period, the calcium signal still peaks during periods of no swimming, indicating that this is a motor-related neuron instead of a visually driven neuron ($|CCFM| = -0.40| > CCFF = 0.18$). It might be involved in inhibiting motor output, or in suppressing behaviours that should not be executed during vigorous swimming. Not all motor-off neurons had this property; some were more visually driven (Fig. 4g).

Motor-off units were concentrated in the dorsal hindbrain (Fig. 5d), in the cerebellum, in the inferior olive, in the ventral midbrain near the NMLF and the pretectum, and in the habenula and pallidum.

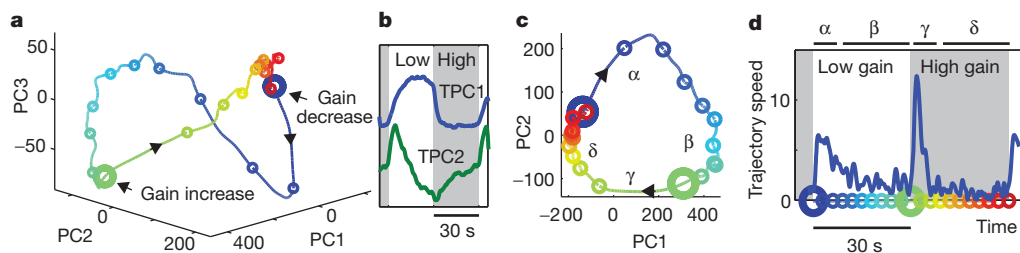


Figure 3 | Low-dimensional representation of neural network dynamics. **a**, Projection of activity of all detected units from all fish onto the first three principal components (PC1, PC2 and PC3) derived from principal components analysis. Circles, intervals of three seconds. **b**, First two temporal principal components (TPCs). TPC1 shows elevated activity during low gain periods; TPC2 shows transient activity after a gain decrease followed by a slow dip (see

Supplementary Figs 20 and 21). **c**, Top view of **a** (α , transient dynamics after switch to low gain; β , steady state during low gain; γ , transient dynamics after switch to high gain; δ , steady state during high gain). **d**, Speed through phase space over one low-high gain period, showing accelerated trajectory speed after gain changes. Colours in **a**–**d** represent time points as shown on the *x* axis of **d**.

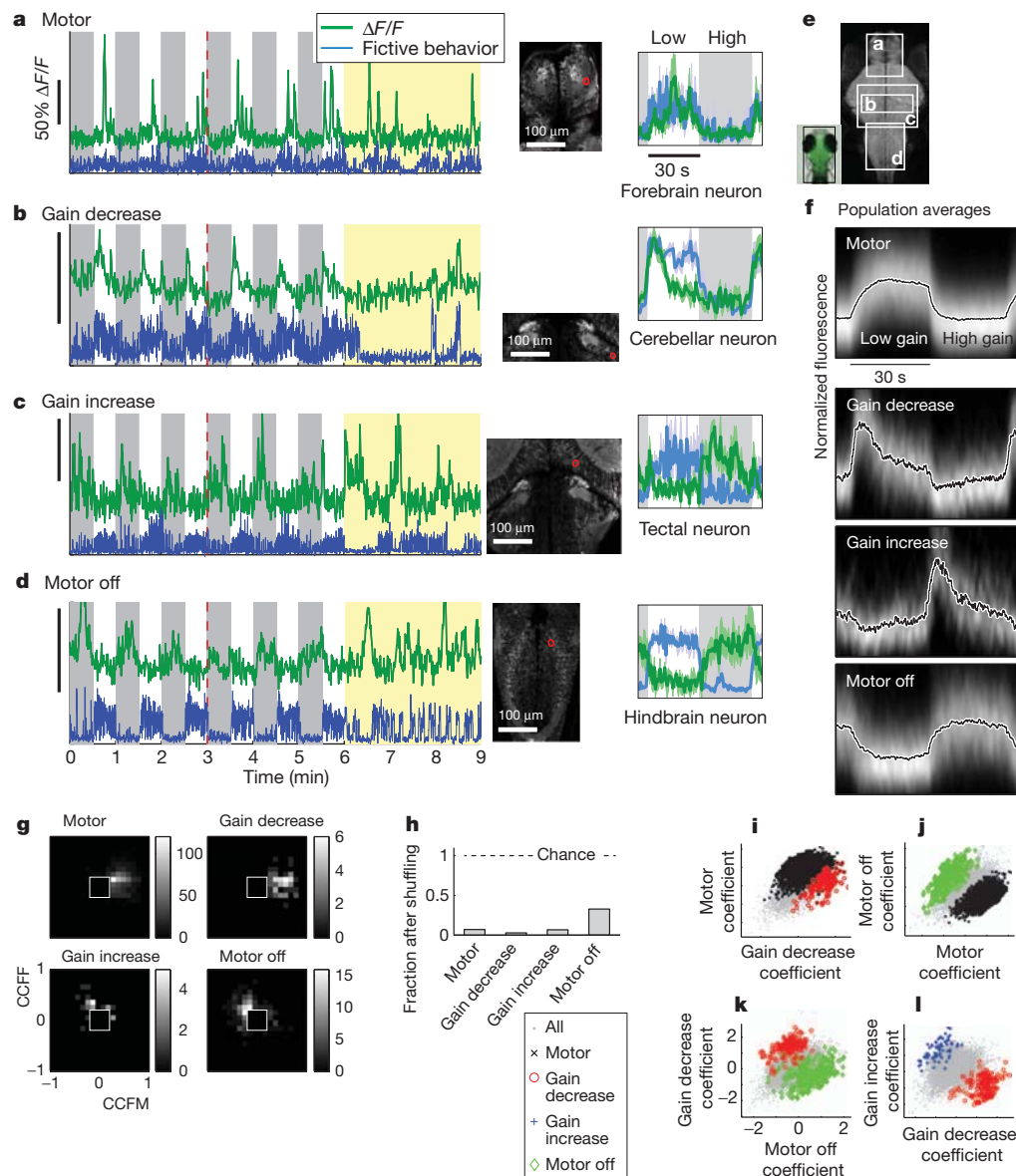


Figure 4 | Four types of neural dynamics during adaptive motor control. **a–d**, Motor-related activity in a single neuron in the forebrain (**a**), transient activity after decreases in gain in a neuron in the deep cerebellum (**b**), transient activity after increases in gain in a visually driven neuron in the tectum (**c**), and motor-off activity in a neuron in the dorsal hindbrain (**d**). Left panels: white, low gain periods; grey, high gain periods; yellow, stimulus replay period from the previous 3 minutes (the start of the replayed portion is indicated by the dashed red line); blue, fictive swim signals; green, single-neuron fluorescence signal. Scale bars, $50\% \Delta F/F$. Middle panels, imaged plane (the neuron is outlined in red). Right panels, fluorescence traces (green) and behaviours (blue), averaged over six gain repetitions, with standard errors (light green and light blue). **e**, Anatomical locations of recordings in **a–d**. **f**, Population data for neuron types as in **a–d**, with normalized averages (black) overlaid on a heat map of individual traces. **g**, Histograms of CCFM and CCFF. Neurons in the

Correlation-based maps of sites strongly correlating or anticorrelating with locomotion during replay (Fig. 5e, f), as measured using CCFM, are generally consistent with maps of the anatomical locations of ‘motor’ neurons and motor-off neurons (Fig. 5a, d), and reveal additional structures such as two arcs in the dorsal hindbrain.

Population averages summarizing the above four types of neural dynamics are shown in Fig. 4f (individual traces are shown in Supplementary Figs 13–16). The detected units were not false positives resulting from noisy measurements of neural activity, and this is confirmed by shuffling the fluorescence time series, which

motor and gain-down groups are more related to locomotion; neurons in the gain-up group are more responsive to visual input, and the motor-off group is mixed. The empty square in the centre represents values of CCFM and CCFF that are indistinguishable from those arising from noise (see Supplementary Figs 17 and 18). **h**, Control for false positives. When fluorescence traces are scrambled by cutting at 16 random time points and rearranging, the number of detected units falls by a factor of 8.1 on average, indicating that detected units are not a result of false positives (chance level is at 1). **i**, Scatter plot of motor coefficient (average normalized fluorescence during seconds 10–30 of low gain period) versus gain-down coefficient (average fluorescence during seconds 1–8 of low gain period) showing segregation with partial overlap of motor and gain-down units. **j–l**, Similar scatter plots to **i** but for other coefficients. No detected neuron codes for both upwards and downwards gain changes (**l**): transient neuronal activity is specific to the direction of gain change.

causes an eightfold drop in detected units (Fig. 4h, Methods). Figure 4i–l shows that neurons generally only belong to one functional class, but there is some overlap between motor and gain-down neurons (46% of gain-down units also have sustained activity during low gain), and no overlap between gain-down and gain-up units, consistent with previously observed asymmetries between gain adaptation in opposite directions⁴². Supplementary Movie 4 contains anatomical stacks with superimposed clusters of the functionally identified neural classes (same data as Fig. 5). As GCaMP2 has a relatively weak signal-to-noise ratio, and as the brain was sampled a

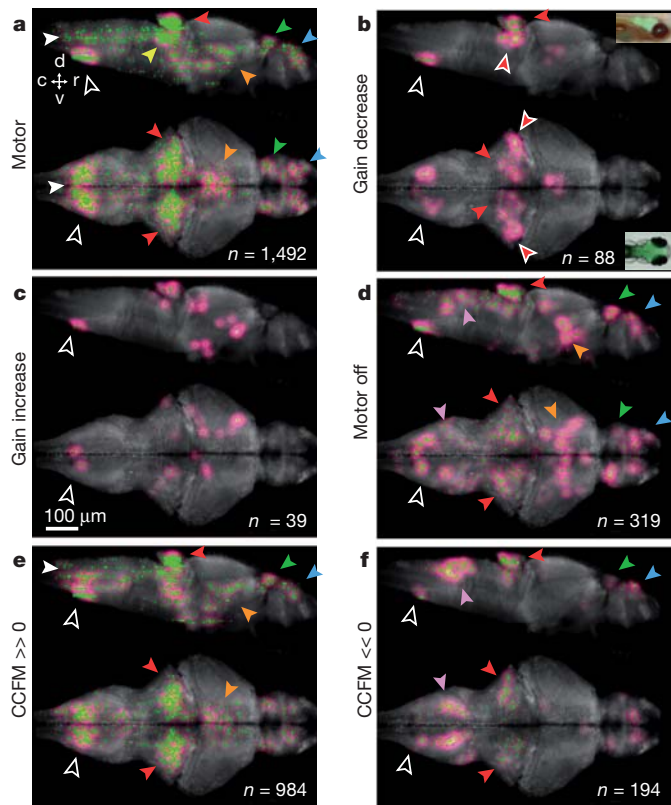


Figure 5 | Anatomical locations of neurons and neuropil regions.

a–d, Anatomical locations are classified by the statistical criteria described in the text (32 fish, normalized by sampling density; see also Supplementary Figs 9–12). **a**, ‘Motor’ neurons. **b**, Gain decrease neurons. **c**, Gain increase neurons. **d**, Motor-off neurons. Green dots, locations of identified sites; magenta, confidence interval scaled by sampling density (Methods). **e**, Sites with activity relating to locomotion as measured during replay by CCFM > 0.4 . **f**, Sites anticorrelating with locomotion, as measured by CCFM < -0.25 . Black arrows with white outline, inferior olive; red arrows, cerebellum; red arrows with white outline, deep lateral cerebellum; yellow arrow, ventral hindbrain nearby cerebellum; white arrows, posterior hindbrain; pink arrows, dorsal posterior hindbrain; orange arrows, NMLF and pretectum area; green arrows, habenula; blue arrows, pallium.

limited number of times, the anatomical maps are not exhaustive, so we quantified the uncertainty in detecting or missing functional units (Supplementary Fig. 10).

Thus, the four types of neural dynamics identified via a dimensionally reduced representation of network activity could be mapped to distinct brain areas. To the best of our knowledge, this is the first brain-wide imaging at the cellular level of activity related to adaptive motor control in a vertebrate brain.

Discussion

The ability to monitor neural activity at single-cell resolution throughout the whole brain of a behaving animal creates new opportunities for studying circuit function during behaviour. The demonstration that paralysed larval zebrafish interact readily with a virtual environment, and the remarkable finding that these animals still showed short-term forms of motor learning in the fictive virtual-reality setup, provided an exciting opportunity to study the circuit dynamics occurring during this behaviour.

Here we identified neural populations activated during specific phases of adaptive locomotion that span multiple areas of the larval zebrafish brain. Both the inferior olive and the cerebellum contained many neurons correlating with adaptive motor control. In mammals, cerebellar circuits play an important role in motor control¹¹ and in fish the cerebellum has been shown to be involved in the selection of

motor programs^{43–45}. Furthermore, the structure of olivocerebellar circuitry in zebrafish is remarkably similar to that of mammals^{36,37,46} (Supplementary Fig. 23a), and the transient gain-down activity observed in the inferior olive and cerebellum may represent error signals driving motor learning mechanisms^{11,40,41,47}.

To test whether the inferior olive is necessary for motor adaptation, we next lesioned it with an infrared laser²⁴. Post lesion, the power of swim bouts in the high- and low-gain settings became statistically indistinguishable (Fig. 6). Although damage to passing axons cannot be ruled out, similar lesions in the dorsal anterior hindbrain did not affect motor adaptation ($P < 0.001$ pre lesion; $P = 0.01$ post lesion; Fig. 6c and Supplementary Fig. 28). The optomotor response was still intact (Supplementary Fig. 24). These results indicate that the inferior olive is necessary for successful adaptation of motor programs to external feedback gain. One possibility is that an error signal is computed in the inferior olive through subtractive interaction of an efference copy of motor output (for example, through inhibitory connections from premotor circuits) and visual feedback from a swim bout (for example, through pretectal projections; see also Supplementary Fig. 27), which then activates appropriate circuits in the cerebellum via climbing fibres. Cerebellar activity may drive changes in motor programs through the deep cerebellum, which in mammals projects to premotor circuits.

Although synaptic plasticity underlies much of motor learning^{11,12}, it is only one of several candidate mechanisms for the behaviour observed here. An alternative idea is that sustained firing rates of neuronal populations, perhaps subsets of the motor and motor-off populations, implement and maintain the different levels of locomotor drive over prolonged periods; for example, through attractor states⁴⁸.

The function of the observed adaptive sensorimotor behaviour may be multifold. On long timescales, development of muscle and body shape will require continuous adjustment of sensorimotor control. On medium timescales, fluctuating body mass due to eating, and fluctuating viscosity of the water require some adaptation of swimming behaviour. In addition, temperature fluctuations cause changes in muscle efficacy⁴⁹ that must be counterbalanced by adjustments of locomotor drive. On short timescales, the relationship between the speed of optic flow following a swim bout depends on the distance to the optical surround (objects farther away induce smaller optic flow), requiring recalibration of the sensorimotor loop on the timescale of seconds. In our experiments, we observed adaptation occurring on such a timescale. Human motor control faces many similar challenges and is subject to continuous recalibration to cope with changing conditions (for example, leg injury, walking on a slippery floor or carrying a heavy bag). Thus, the current study of brain-wide activity during adaptive locomotion is an important step towards

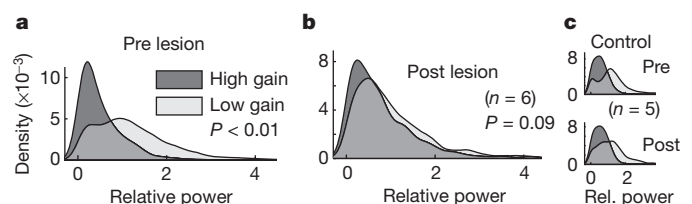


Figure 6 | Effect of inferior olive lesions on visually induced motor adaptation. **a**, Histogram of power per swim bout in the high- and low-gain conditions for six fish, pre lesion. Difference in mean power per fish is significant before lesion ($P < 0.001$, t -test on means of $n = 6$ fish). **b**, Post-lesion histogram. The inferior olive was lesioned in approximately 60 locations with an infrared laser. The fish still performed the optomotor response (Supplementary Fig. 24), but the power of the swim bouts was no longer adjusted significantly to the external feedback gain ($P = 0.09$, same six fish). **c**, Lesions of similar size in the dorsal anterior hindbrain did not impair motor adaptation ($P < 0.001$ pre lesion for $n = 5$ fish, $P = 0.01$ post lesion, Supplementary Figs 23 and 28). These results indicate that the inferior olive is a necessary component of the circuit driving motor adaptation.

understanding entire circuits in the precise context for which they evolved: the flexible control of behaviour in changing environments.

METHODS SUMMARY

Larval zebrafish fictively swam in a one-dimensional virtual environment, while neuronal activity was monitored using a two-photon microscope.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 June 2011; accepted 15 March 2012.

Published online 9 May 2012.

- Wall, P. D., Freeman, J. & Major, D. Dorsal horn cells in spinal and in freely moving rats. *Exp. Neurol.* **19**, 519–529 (1967).
- Flusberg, B. A. *et al.* High-speed, miniaturized fluorescence microscopy in freely moving mice. *Nature Methods* **5**, 935–938 (2008).
- Naumann, E. A., Kampff, A. R., Prober, D. A., Schier, A. F. & Engert, F. Monitoring neural activity with bioluminescence during natural behavior. *Nature Neurosci.* **13**, 513–520 (2010).
- Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L. & Tank, D. W. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neurosci.* **13**, 1433–1440 (2010).
- Maimon, G., Straw, A. D. & Dickinson, M. H. Active flight increases the gain of visual motion processing in *Drosophila*. *Nature Neurosci.* **13**, 393–399 (2010).
- Seelig, J. D. *et al.* Two-photon calcium imaging from head-fixed *Drosophila* during optomotor walking behavior. *Nature Methods* **7**, 535–540 (2010).
- Fry, S. N., Rohrseitz, N., Straw, A. D. & Dickinson, M. H. Visual control of flight speed in *Drosophila melanogaster*. *J. Exp. Biol.* **212**, 1120–1130 (2009).
- Möhl, B. Short-term learning during flight control in *Locusta migratoria*. *J. Comp. Physiol.* **163**, 803–812 (1988).
- Wolf, R., Voss, A., Hein, S., Heisenberg, M. & Sullivan, G. D. Can a fly ride a bicycle? *Phil. Trans. R. Soc. Lond. B* **337**, 261–269 (1992).
- du Lac, S., Raymond, J. L., Sejnowski, T. J. & Lisberger, S. G. Learning and memory in the vestibulo-ocular reflex. *Annu. Rev. Neurosci.* **18**, 409–441 (1995).
- Raymond, J. L., Lisberger, S. G. & Mauk, M. D. The cerebellum: a neuronal learning machine? *Science* **272**, 1126–1131 (1996).
- Gilbert, P. F. & Thach, W. T. Purkinje cell activity during motor learning. *Brain Res.* **128**, 309–328 (1977).
- Körding, K. P. & Wolpert, D. M. Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247 (2004).
- Portugues, R. & Engert, F. Adaptive locomotor behavior in larval zebrafish. *Front. Syst. Neurosci.* **5**, 72 (2011).
- Rock, I. & Smith, D. The optomotor response and induced motion of the self. *Perception* **15**, 497–502 (1986).
- Orger, M. B., Smear, M. C., Anstis, S. M. & Baier, H. Perception of Fourier and non-Fourier motion by larval zebrafish. *Nature Neurosci.* **3**, 1128–1133 (2000).
- Gahtan, E., Sankrithi, N., Campos, J. B. & O'Malley, D. M. Evidence for a widespread brain stem escape network in larval zebrafish. *J. Neurophysiol.* **87**, 608–614 (2002).
- Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
- Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
- Higashijima, S.-I., Masino, M. A., Mandel, G. & Fetcho, J. R. Imaging neuronal activity during zebrafish behavior with a genetically encoded calcium indicator. *J. Neurophysiol.* **90**, 3986–3997 (2003).
- Bene, F. D. *et al.* Filtering of visual information in the tectum by an identified neural circuit. *Science* **330**, 669–673 (2010).
- Douglass, A. D., Kraves, S., Deisseroth, K., Schier, A. F. & Engert, F. Escape behavior elicited by single, channelrhodopsin-2-evoked spikes in zebrafish somatosensory neurons. *Curr. Biol.* **18**, 1133–1137 (2008).
- Chong, M. & Drapeau, P. Interaction between hindbrain and spinal networks during the development of locomotion in zebrafish. *Dev. Neurobiol.* **67**, 933–947 (2007).
- Orger, M. B., Kampff, A. R., Severi, K. E., Bollmann, J. H. & Engert, F. Control of visually guided behavior by distinct populations of spinal projection neurons. *Nature Neurosci.* **11**, 327–333 (2008).
- O'Malley, D. M., Kao, Y. H. & Fetcho, J. R. Imaging the functional organization of zebrafish hindbrain segments during escape behaviors. *Neuron* **17**, 1145–1155 (1996).
- Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
- Masino, M. A. & Fetcho, J. R. Fictive swimming motor patterns in wild type and mutant larval zebrafish. *J. Neurophysiol.* **93**, 3177–3188 (2005).
- Cohen, A. H. & Wallén, P. The neuronal correlate of locomotion in fish. “fictive swimming” induced in an *in vitro* preparation of the lamprey spinal cord. *Exp. Brain Res.* **41**, 11–18 (1980).
- Tallini, Y. N. *et al.* Imaging cellular signals in the heart *in vivo*: cardiac expression of the high-signal Ca^{2+} indicator GCaMP2. *Proc. Natl Acad. Sci. USA* **103**, 4753–4758 (2006).
- Park, H. C. *et al.* Analysis of upstream elements in the *HuC* promoter leads to the establishment of transgenic zebrafish with fluorescent neurons. *Dev. Biol.* **227**, 279–293 (2000).
- Ito, M., Shiida, T., Yagi, N. & Yamamoto, M. Visual influence on rabbit horizontal vestibulo-ocular reflex presumably effected via the cerebellar flocculus. *Brain Res.* **65**, 170–174 (1974).
- Mazor, O. & Laurent, G. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron* **48**, 661–673 (2005).
- Yaksi, E., von Saint Paul, F., Niessing, J., Bunschuh, S. T. & Friedrich, R. W. Transformation of odor representations in target areas of the olfactory bulb. *Nature Neurosci.* **12**, 474–482 (2009).
- Kinkhabwala, A. *et al.* A structural and functional ground plan for neurons in the hindbrain of zebrafish. *Proc. Natl Acad. Sci. USA* **108**, 1164–1169 (2011).
- Koyama, M., Kinkhabwala, A., Satou, C., Higashijima, S.-I. & Fetcho, J. Mapping a sensory-motor network onto a structural and functional ground plan in the hindbrain. *Proc. Natl Acad. Sci. USA* **108**, 1170–1175 (2011).
- Bae, Y.-K. *et al.* Anatomy of zebrafish cerebellum and screen for mutations affecting its development. *Dev. Biol.* **330**, 406–426 (2009).
- Kani, S. *et al.* Proneural gene-linked neurogenesis in zebrafish cerebellum. *Dev. Biol.* **343**, 1–17 (2010).
- Volkman, K., Chen, Y.-Y., Harris, M. P., Wullmann, M. F. & Köster, R. W. The zebrafish cerebellar upper rhombic lip generates tegmental hindbrain nuclei by long-distance migration in an evolutionary conserved manner. *J. Comp. Neurol.* **518**, 2794–2817 (2010).
- Wullmann, M. F., Rupp, B. & Reichert, H. *Neuroanatomy of the Zebrafish Brain: a Topological Atlas* (Birkhäuser, 1996).
- Marr, D. A theory of cerebellar cortex. *J. Physiol.* **202**, 437–470 (1969).
- Albus, J. A theory of cerebellar function. *Math. Biosci.* **10**, 25–61 (1971).
- Boyden, E. S. & Raymond, J. L. Active reversal of motor memories reveals rules governing memory encoding. *Neuron* **39**, 1031–1042 (2003).
- Matsumoto, N., Yoshida, M. & Uematsu, K. Effects of partial ablation of the cerebellum on sustained swimming in goldfish. *Brain Behav. Evol.* **70**, 105–114 (2007).
- Roberts, B. L., van Rossem, A. & de Jager, S. The influence of cerebellar lesions on the swimming performance of the trout. *J. Exp. Biol.* **167**, 171–178 (1992).
- Aizenberg, M. & Schuman, E. M. Cerebellar-dependent learning in larval zebrafish. *J. Neurosci.* **31**, 8708–8712 (2011).
- Ma, L., Punnamootil, B., Rinkwitz, S. & Baker, R. Mosaic *hoxb4a* neuronal pleiotropism in zebrafish caudal hindbrain. *PLoS ONE* **4**, e5944 (2009).
- De Zeeuw, C. I. *et al.* Microcircuitry and function of the inferior olive. *Trends Neurosci.* **21**, 391–400 (1998).
- Miri, A. *et al.* Spatial gradients and multidimensional dynamics in a neural integrator circuit. *Nature Neurosci.* **14**, 1150–1159 (2011).
- Bennett, A. F. Temperature and muscle. *J. Exp. Biol.* **115**, 333–344 (1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to D. Schoppik for teaching M.B.A. the fictive swimming preparation, to K.-H. Huang for carrying out spinal calcium green injections, and to M. Concha, R. Baker and L.-H. Ma for advice on anatomy. We thank M. Meister, B. Ölveczky, D. Wolpert, E. Mukamel, M. Yartsev, D. Schoppik, D. Hildebrand, E. Naumann, A. Kampff, P. Latham, T. Dunn and members of the Engert laboratory for useful discussions and comments on the manuscript. We thank P. Oteiza and R. Hellmich for help with anatomy and figures, and A. Viel for use of laboratory space. M.B.A. thanks D. Wolpert and E. Santos for support. This work was supported by a Sir Henry Wellcome Fellowship from the Wellcome Trust (M.B.A.), a K99 grant no. 5K99NS062780-2 (M.B.O.) and National Institutes of Health grants 5R01EY014429 and RC2NS069407 (F.E.).

Author Contributions M.B.A. developed the fictive virtual-reality paradigm, carried out the experiments, analysed the data, and built the setup and software. M.B.A., F.E. and R.P. conceived the experiments. M.B.O., D.N.R., J.M.L. and A.F.S. generated the transgenic *elavl3:GCaMP2* fish line. M.B.O. generated the transgenic alpha tubulin:C3PA-GFP fish line. All authors discussed the data and the manuscript. M.B.A. wrote the manuscript with the assistance of R.P., M.B.O. and F.E.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.E. (florian@mcb.harvard.edu).

METHODS

Fictively driven virtual environment. Wild-type larvae of the WIK strain were used for behavioural experiments. For reticulospinal imaging, nacre fish on a WIK background were used, and for two-photon imaging, nacre fish expressing GCaMP2 under control of the *elavl3* promoter (refs 20, 30) were used, again on a WIK background. All experiments were approved by the Standing Committee on the Use of Animals in Research and Training of Harvard University. Zebrafish larvae of 6 or 7 d.p.f. were briefly anaesthetized with MS222 and paralysed by injection with a 1 mg ml⁻¹ bungarotoxin solution (Sigma-Aldrich), then suspended from structural pipettes (Supplementary Fig. 3) or embedded in agarose but with the agarose around the tail removed. Motor nerve recordings were made with a Multiclamp 700B amplifier, simultaneously with two-photon imaging. Experiments were done at room temperature in filtered facility fish water. Visual scenes were projected onto a diffusive screen underneath the Petri dish containing the fish, using a mini projector whose light source was replaced by a red Luxeon Rebel light-emitting diode (LED) that was pulsed in synchrony with the fast scan mirror, so that illumination only occurred at the edges of the image where the scan mirror changed direction (typically at 800 Hz) to avoid any corruption of the two-photon images. Visual scenes consisted of square gratings with spatial period 12 mm moving at 1 cm s⁻¹ from tail to head in the absence of motor nerve signals (−1 cm s⁻¹). When the processed swim signal was above an automatically set threshold (see Supplementary Methods and Supplementary Fig. 2), the locomotor drive was defined as the area underneath the curve of the processed swim signal during the current and the previous video frame. The processed swim signal was defined as the standard deviation of the raw swim signal in a sliding window of 15 ms (see Fig. 1d). In the presence of such motor nerve signals, the instantaneous virtual-fish velocity was set, 60 times per second (at the rate of the 60-Hz projector), to −1 cm s⁻¹ + (gain × instantaneous locomotor drive), with the gain set experimentally, after which the velocity decayed back to −1 cm s⁻¹ at a rate of −15 cm s⁻², approximately matched to freely swimming fish dynamics (Supplementary Fig. 1). The high gain was chosen to be two to five times higher than the low gain, and these values bracketed the ‘natural’ gain setting that described the transformation of motor activity into optic flow in a freely swimming fish. The high- and low-gain settings were manually adjusted for each fish, as different fish showed different ranges of adaptability. Some fish showed a transient increase in fictive motor output followed by a decrease after a gain-down change; these fish were discarded from the gain-down data set because transient neural activity could not be distinguished from motor-related activity (rejection criterion: $P < 0.03$, paired *t*-test on fictive signal averages over seconds 0–15 versus averages over seconds 15–30 after gain-down change).

Generation of transgenic fish. To generate the transgenic line Tg(-1.7CaTuba:C3PA-GFP) the C3PA-GFP open reading frame and polyadenylation sequence⁵⁰ was PCR amplified using the primers cttttgctttttctc acagGTGAGCAAGGGCGAGGAGC (forward, coding sequence upper case) and GGggtatcTGGACAAACCACAACCTAG (reverse) and fused in a second PCR step to the goldfish α 1-tubulin *cis*-regulatory sequence⁵¹, and first exon and intron amplified with the forward primer ggACGCGTgctcccgactcagatc and complementary reverse primer GCTCCTCGCCCTTGCTCActgtgaagaaa aggcaaaag, resulting in an in frame replacement of the α -tubulin coding sequence with C3PA-GFP. This cassette was cloned into a plasmid with Tol2 arms using restriction enzymes Mlu1 and BamH1. The transgenic line Tg(*elavl3*:GCaMP2) was made by first cloning GCaMP2 downstream of an AttR1-R2 cassette flanked by Tol2 arms, and then placing it under the control of *elavl3* (*Hu*C) *cis*-regulatory sequences including the first exon with the start codon mutated, and first intron²⁰ via L/R recombination with an AttL flanked *elavl3* entry clone. The resulting plasmids were injected into 1-cell-stage embryos at a concentration of 20 ng μ l⁻¹ in combination with Tol2 transposase RNA at a concentration of 25 ng μ l⁻¹. Founders were selected based on high and spatially uniform expression.

Two-photon microscopy. A custom-built laser scanning two-photon microscope, using a Spectra Physics Mai Tai pulsed infrared laser, was used to monitor fluorescence in the brain. A photomultiplier tube (PMT) with a green bandpass filter and a laser-blocking filter was used to monitor green fluorescence. The fish was illuminated from below by a near-infrared LED light source whose light path was combined with the light path of the projector using a dichroic mirror (a ‘hot mirror’, Edmund Optics; not shown in Fig. 1). The image of the fish was registered above the objective using another dichroic mirror by a camera (AVT Stingray). Areas of the brain were scanned at 1.5 to 3 Hz, typically with 4–8 μ m between imaging planes. Fish typically showed robust behaviour for a period of 3–5 h, so that not the entire brain of every fish could be scanned (although almost 100% coverage was possible in some fish). Thirty-two fish were used to non-homogeneously cover six times the volume of the brain (see Supplementary Fig. 9). The electrophysiology, stimulus presentation and the two-photon microscope were controlled by a single piece of

software custom written in C# (Microsoft). Two data acquisition cards (National Instruments) were used to acquire imaging and electrophysiology data, and were synchronized using periodic digital pulses.

Analysis of the two-photon images. Image analysis software was written in Matlab (Mathworks). We developed a novel method for automatically extracting regions of interest from two-photon movies that contain many neurons. A square ROI, half the size of a neuron, is swept over all locations of the imaging plane. At each location, a fluorescence time series, averaged over the ROI, is extracted and converted to a statistic for the ‘peaky-ness’ of the fluorescence signal at that point. If $f_{x,y}(t)$ is the fluorescence of the pixel at x, y at time t , and $\bar{f}_{x,y}$ is the average fluorescence at pixel x, y , and \bar{f} is the average of $\bar{f}_{x,y}$, then the statistic is defined as:

$$m_{x,y} = \left\langle \left(\frac{f_{x,y}(t) - \bar{f}_{x,y}}{\bar{f}_{x,y} + \bar{f}} \right)^3 \right\rangle_{ROI, t}$$

where $\langle \dots \rangle_{ROI}$ denotes the spatial average over the ROI and $\langle \dots \rangle_t$ means average over time. This measure was chosen because it bears a resemblance to the usual $\Delta F/F$ but contains an offset to counteract the undesirable amplification of noise in areas of low fluorescence; the third power was chosen because it nonlinearly converts peaks in the fluorescence signal to larger values of the statistic. This measure—one of several tested—yielded spatial signal maps $m_{x,y}$, which tended to be at least as sensitive as sets of ROIs selected manually from observation of the raw and $\Delta F/F$ movies. In general, the maps were also more complete than ROIs obtained using PCA and independent component analysis (ICA) methods⁵⁰, which detected large events in which many neurons participated, but often missed small, sparse activation of single neurons. Thus, a spatial map of activity is obtained as in the top panel of Fig. 2c. This map is smoothed by convolution with a Gaussian function the size of a neuron, and the maxima are designated as points of interest. Points that were less than a cell-body length away from one another, including ‘chains’ of such points, which sometimes occurred in regions of activated neuropil, were pruned to a single point at the centre using an automated procedure. These points of interest formed the centre of new ROIs and were found to be positioned at the centre of neurons in the majority of cases (for example, Fig. 2c). The fluorescence time series of these ROIs formed the basis of the functional classification. The time series could be correlated with the time series of all pixels in a region around the point of interest, leading to a spatial map of correlation coefficients (Fig. 2d, top panel). This method extracted shapes of single neurons because pixel values within a single neuron tend to correlate highly over time. In many cases, the correlational maps also revealed other neurons whose activity was highly correlated with the neuron at the point of interest (Supplementary Fig. 7).

Automatic localization of imaging planes in a reference brain. An entire 6 d.p.f. larval brain (Tg(*elavl3*:GCaMP2) fish treated with 1-phenyl 2-thiourea (PTU) to inhibit eye pigmentation) was scanned and used as a reference brain. Individual imaged planes of other fish were registered onto the reference brain as follows (software was written in Matlab). The reference stack was intensity-normalized per z plane, and thresholded to dampen the effect of brightly labelled neurons on the registration. The peak of the cross-correlation between the mean-subtracted z -planes reference stack and a mean-subtracted image was calculated for each z plane of the reference stack, giving the most likely x and y positions for that z plane (Fig. 2b, Supplementary Fig. 8 and Supplementary Movie 3). The z location was chosen to be the peak of these correlation values, which fixed the x, y and z positions in the entire stack. Manual inspection showed that 82% of images were correctly registered; the remaining images were localized manually. To estimate the accuracy of localization, the reference brain (A) was used as well as a second brain (B). A plane (250 by 270 μ m) was randomly selected from brain A, and mapped to brain B by the algorithm. Next, the corresponding plane in brain B was mapped back to A. This was repeated 500 times. Thus, the original location could be compared to the location after two mappings by the algorithm; 500 discrepancies were obtained in this way. The uncertainty in the mapping algorithm was then approximated by the standard deviation of these 500 discrepancies, divided by $\sqrt{2}$ (because two errors accumulate by the double mapping). This was 19.5 μ m, so we assumed the slightly higher value of 25 μ m for the precision of the algorithm. In 14% of cases, a gross error occurred of greater than 60 μ m. These were easily spotted by eye (compare to the manual mapping of our data after gross discrepancies; see above). The unidirectional gross error rate was therefore estimated to be 7%.

As each fish contributed a partial brain to the data set, a total volume of six times the volume of the brain was obtained (assuming a 4- μ m z resolution), with certain areas covered more densely than others (see Supplementary Figs 9 and 10). The density maps of Fig. 5 were normalized by the sampling density so that maps were not biased towards the most densely sampled areas. Roughly 1% of all neurons were classified as being active by our activity detection algorithm. Of these, under our statistical criteria, about 20% could be related to the behaviour by

classification into the four groups described in the text. All major clusters in Fig. 5 appeared in more than one fish; neurons were removed if, within a 50- μm sphere, no other neuron within the same class from a different fish was present.

Analysis of neuronal activity. After extracting fluorescence time series from the ROIs (Methods), the traces were analysed in more detail. To distinguish neural activity driven by visual input from activity relating to locomotion, we first smoothed and combined the left and right fictive channel recordings to generate a single quantitative descriptor of motor output, M . The motor output and the fluorescence trace were cross-correlated during the replay period (see text) to yield a fluorescence-motor correlation coefficient CCFM. The correlation coefficient of the fluorescence trace during and before replay was termed CCFF. High values of CCFM indicate neural activity relating to locomotion, and high values of CCFF indicate visually driven activity (because the stimulus during replay is a repetition of the preceding stimulus; visually driven neurons should therefore respond in a similar way). In the CCFF versus CCFM histograms (Fig. 4g), there is a region in the centre that cannot be distinguished from noise, which was derived from a fish that did not receive visual input and that did not swim (Supplementary Figs 17 and 18), and which is therefore left blank in the figures.

For dimensionality reduction, principal components of the matrix containing the fluorescence traces of all detected sites (size $n_{\text{sites}} \times n_{\text{time points}}$) were found using the princomp function in Matlab, and consisted of vectors of length n_{sites} . Activity traces from all detected sites, averaged over the six low-high gain repetitions, were projected (by the dot product) onto the first three of these principal components, to obtain three time series (length $n_{\text{time points}}$). These time series served as a 'summary' of activity of all sites across brain regions and fish, as they can be used to approximately reconstruct the activity of the detected sites (using the principal components). The three time series were then visualized by plotting them on three axes, rendering a three-dimensional curve as shown in Fig. 3.

The motor-, gain decrease-, gain increase- and motor-off activity patterns were detected as described in the text. As a large number of units ($n = 9,814$ over 32 fish) contributed to the data set, a concern is that the detected units are a result of false positives, arising from randomly fluctuating signals being classified purely by chance in one of the four functional categories. To address this, we applied a shuffle test, in which fluorescence traces were cut at 16 random time points and randomly rearranged, after which analysis proceeded as normal. The fact that the number of units in the four categories fell by a factor of 8.1 (Fig. 4h) indicates that the detected units did not arise by chance.

The motor-, gain increase-, gain decrease- and motor-off-based description of neural activity (Fig. 5a–d) is complementary to the CCFM- and CCFF-based description (Fig. 5e–f); the motor-based description derives from the gain-adaptation assay, and the CCFM- and CCFF-based description derives from the replay period, with both being useful functional descriptors of neural activity.

Lesions. Inferior olive lesions were performed by pre-selecting about 60 sites in an averaged two-photon image, then shining an infrared laser on them at 850 nm, at 900 mW power outside the cage. The pulses lasted 200 ms per site. During the exposure, the laser beam spiralled over a circle of 1 μm (ref. 24). Brief but large increases in emitted light intensity indicated a successful lesion; in the absence of such signals, the site was exposed once more, and abandoned if the second attempt failed.

50. Ruta, V., Datta, S. R., Vasconcelos, M. L., Freeland, J., Looger, L. L. & Axel, R. A dimorphic pheromone circuit in *Drosophila* from sensory input to descending output. *Nature* **468**, 686–690 (2010).
51. Hieber, V., Dai, X., Foreman, M. & Goldman, D. Induction of $\alpha 1$ -tubulin gene expression during development and regeneration of the fish central nervous system. *J. Neurobiol.* **37**, 429–440 (1998).
52. Mukamel, E. A., Nimmerjahn, A. & Schnitzer, M. J. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* **63**, 747–760 (2009).

Statistical geochemistry reveals disruption in secular lithospheric evolution about 2.5 Gyr ago

C. Brenhin Keller¹ & Blair Schoene¹

The Earth has cooled over the past 4.5 billion years (Gyr) as a result of surface heat loss and declining radiogenic heat production. Igneous geochemistry has been used to understand how changing heat flux influenced Archaean geodynamics^{1,2}, but records of systematic geochemical evolution are complicated by heterogeneity of the rock record and uncertainties regarding selection and preservation bias^{3–5}. Here we apply statistical sampling techniques to a geochemical database of about 70,000 samples from the continental igneous rock record to produce a comprehensive record of secular geochemical evolution throughout Earth history. Consistent with secular mantle cooling, compatible and incompatible elements in basalts record gradually decreasing mantle melt fraction through time. Superimposed on this gradual evolution is a pervasive geochemical discontinuity occurring about 2.5 Gyr ago, involving substantial decreases in mantle melt fraction in basalts, and in indicators of deep crustal melting and fractionation, such as Na/K, Eu/Eu* (europium anomaly⁴) and La/Yb ratios in felsic rocks. Along with an increase in preserved crustal thickness across the Archaean/Proterozoic boundary^{6,7}, these data are consistent with a model in which high-degree Archaean mantle melting produced a thick, mafic lower crust and consequent deep crustal delamination and melting—leading to abundant tonalite–trondhjemite–granodiorite magmatism and a thin preserved Archaean crust. The coincidence of the observed changes in geochemistry and crustal thickness with stepwise atmospheric oxidation⁸ at the end of the Archaean eon provides a significant temporal link between deep Earth geochemical processes and the rise of atmospheric oxygen on the Earth.

Although it is generally assumed that mantle potential temperatures were higher in the Archaean, the ratio of terrestrial heat production to heat loss through time is very poorly constrained⁹. The resulting lack of confidence in models of secular thermal evolution has contributed to uncertainty regarding the style of Archaean geodynamics, given the temperature dependence of rock deformation mechanisms, mantle and crustal melting, and magma geochemistry¹. The existence of high-degree-melt rocks such as komatiites, the predominance of the high-sodium tonalite–trondhjemite–granodiorite (TTG) series, and the purported structural evolution of granite–greenstone terranes have all been used to suggest that plate tectonics may have been different in style or even nonexistent during the Archaean⁹. However, thermobarometric estimates of crustal geotherms similar to those for modern continents, geochronological evidence for tectonic-scale underthrusting, and the reported occurrence of subduction signatures such as high-field-strength element depletion have led other authors to argue that plate tectonics has existed in its present state since at least 3.0 or as early as 3.8 Gyr ago^{9,10}.

Much like those generated in modern tectonic environments, rocks from preserved Archaean crust are characterized by extreme geochemical heterogeneity. To obtain systematic trends from this record, studies of igneous geochemistry over Earth history have typically relied on small, carefully selected data sets from well characterized terranes^{3,11,12}. We take a different approach, reasoning that a sufficiently large, representative data set can be used to define global geochemical averages

very precisely; and that changes in this average over time should reflect the influence of long-term geodynamic drivers such as mantle temperature.

To this end, we have constructed a database of ~70,000 major and trace element whole-rock analyses from pre-existing sources^{3,4,11,13} that report sample location as well as crystallization age and uncertainty. These data are superimposed on existing geophysical models (Crust 2.0¹⁴ and TC1⁵) that estimate present-day crustal and lithospheric parameters. Results for global mean geochemistry with time are generated by Monte Carlo analysis with weighted bootstrap resampling (Methods; Supplementary Fig. 1). Resampling weights are inversely related to spatiotemporal sample density, to achieve a maximally uniform posterior sample density distribution and minimize selection bias (Supplementary Fig. 2). The result is a best estimate of the average igneous geochemistry of exposed continental crust through time, reported as means with associated 2-standard-error (95% confidence interval) uncertainties of the mean for 100-Myr intervals between 0 and 3.8 Gyr. In order to distinguish between primary mantle melting and subsequent evolution and fractionation processes, we extract results from two silica ranges: 62–74% SiO₂ for intermediate to felsic samples, and 43–51% SiO₂ for less fractionated mafic samples.

The results for the least fractionated (43–51% SiO₂) mafic samples preserved in the continental igneous rock record reveal a systematic decrease in compatible element concentration and a corresponding increase in incompatible element concentration through time (Fig. 1). We can show that these trends reflect decreases in mantle partial melting through time, as a consequence of secular mantle cooling; alternative explanations such as a change in the degree of fractionation through time are incompatible with the observed data (Methods; Supplementary Figs 8, 9). Although such low-silica samples do not generally represent primitive melts, compositional trends in these samples nonetheless reflect source processes, given a representative data set, allowing us to estimate minimum percent mantle partial melting through time (Methods). In order to derive such minimum estimates, we applied least-squares optimization to match thermodynamic batch melting simulations of primitive mantle from MELTS^{15,16} with Monte Carlo simulations of basaltic major-element concentrations through time (Methods). The results (Fig. 1g) show a systematic decrease in mantle melt fraction from ~0.35 in the Archaean to ~0.1 today. These results agree well with studies of mantle xenoliths which suggest that Archaean subcontinental mantle lithosphere was produced by extraction of ~40% melt¹⁷.

In addition to the gradual trend expected from secular cooling, our results show a sudden, systematic decrease of ~10 percentage points (from ~27% to ~17%) in the average degree of mantle melting at ~2.5 Gyr (Fig. 1g). Due to the Earth's thermal inertia, it is unlikely that this event is the result of a rapid change in mantle potential temperature. Although a rapid change in the style of tectonics and crust formation at ~2.5 Gyr could provide one possible explanation for the rapid decrease in average mantle melting, an abrupt response in degree of melting to a gradual cooling path may provide a more parsimonious explanation; both experiments and thermodynamic

¹Department of Geosciences, Princeton University, Guyot Hall, Washington Road, Princeton, New Jersey 08544, USA.

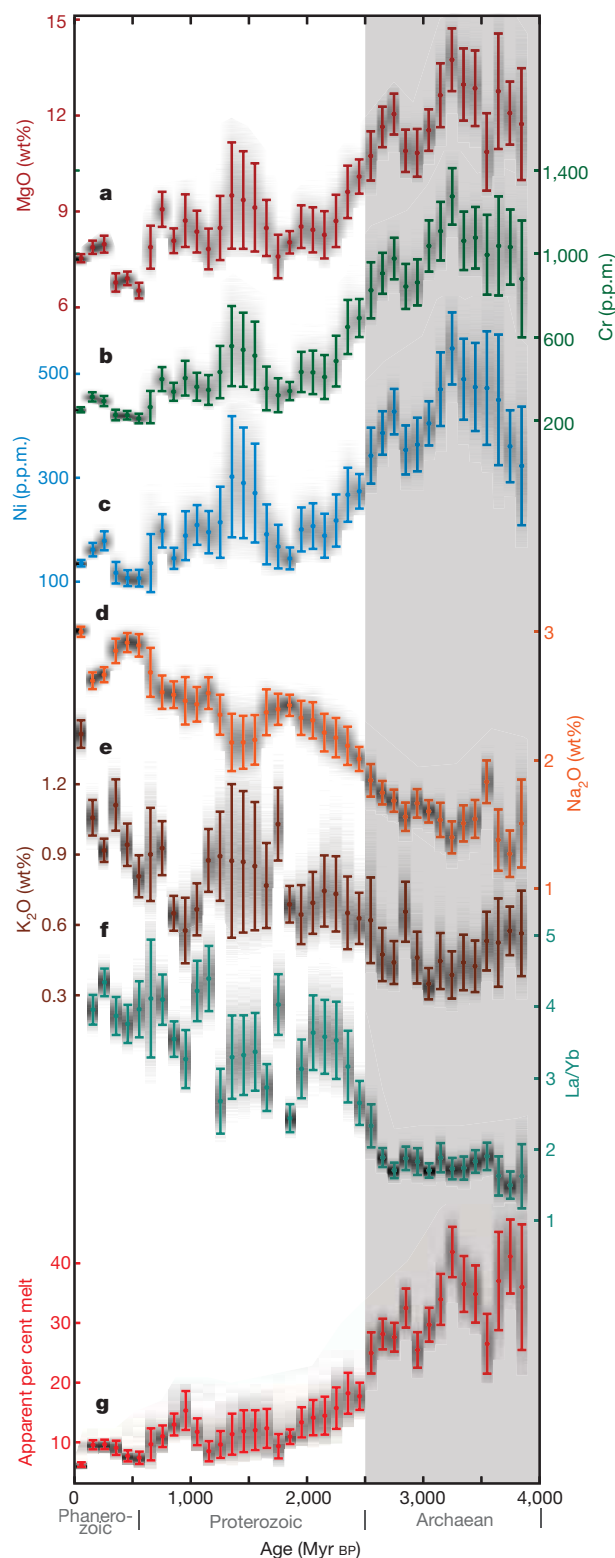


Figure 1 | Secular compositional evolution of mafic lithologies. Secular mantle cooling is reflected in trends over time in the concentrations of compatible (a–c) and incompatible (d, e) elements, and La/Yb ratio (f) in relatively unfractionated rocks (43–51% SiO₂); and in the apparent degree of mantle melting (g), inferred from major-element geochemistry. Error bars show 2-s.e.m. (standard error of the mean) uncertainties; shading illustrates density of the individual Monte Carlo simulations.

models indicate substantially nonlinear mantle melting as a function of temperature, with changes in slope corresponding to clinopyroxene exhaustion¹⁸. Trace-element geochemistry provides some additional

insight: simple trace-element modelling (Methods) suggests that the marked increases in La/Yb (Fig. 1f), La/Sm and Sm/Yb cannot be explained by the trend in degree of mantle melting in Fig. 1g alone, indicating a change in either trace-element composition of the source or effective bulk partition coefficients.

Geochemical trends in high-silica (62–74% SiO₂) rocks parallel many of those observed in basalts. Sodium, however, represents a notable exception. The high Na/K (Fig. 2a) and Na/(Ca + Na + K) of the TTG series have pervaded the discussion of Archaean crust formation^{3,4,11,19–21}. These geochemical signatures, along with high ratios of light to heavy rare-earth elements (for example, La/Yb; Fig. 2b), have been used to argue for the formation of TTG magmas by high-pressure (>40 km depth) partial melting of metabasalt at pressures above the Na-plagioclase stability field in the presence of restitic amphibole and/or garnet^{3,4,11,19–21}. Experimental data for partial melting of mafic rocks at mid- to lower-crustal pressures and temperatures have demonstrated a positive correlation between pressure and sodium concentration in the melt²⁰. Trace-element results from our data set show correlations predicted by such a model: the secular evolution of Eu/Eu* and Sr (Fig. 2c,d) parallels that of Na/K in felsic rocks, arguing that deep crustal melting (and/or high-pressure fractional crystallization²²) played a larger role in forming preserved Archaean crust than post-Archaean crust.

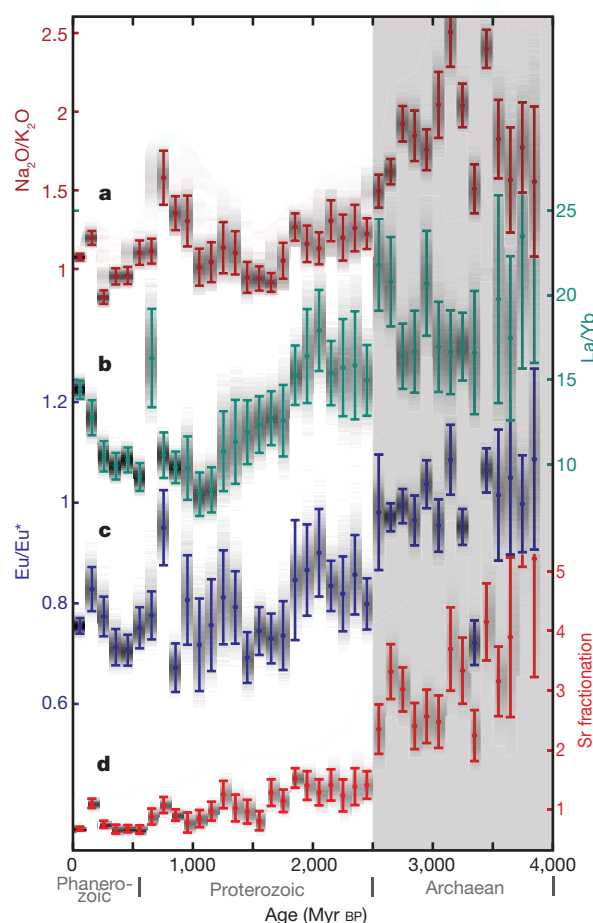


Figure 2 | Secular compositional evolution of felsic lithologies. A shift in lithospheric evolution at ~2.5 Gyr is suggested by trends in the major-element (a) and trace-element (b, c) compositions of felsic rocks (62–73% SiO₂), and in the relative abundance of Sr in these rocks (d), normalized to Sr in mafic (43–51% SiO₂) samples. As Na, Eu and Sr are all partitioned into plagioclase, higher abundance of these elements corresponds to decreased plagioclase fractionation. Error bars show 2-s.e.m. uncertainties; shading illustrates density of the individual Monte Carlo simulations.

The two most frequently proposed settings for such deep crustal anatexis and TTG generation are the melting of subducted oceanic crust; and the melting of overthickened mafic lower crust, which derives its basalt from mantle melting and crustal underplating^{19,21}. Of these, the latter provides a potential genetic link between TTG generation and observed high-degree mantle melting in the Archaean; extensive basalt production could lead to a thicker and/or more mafic lower continental crust, which would more often result in TTG generation. However, any process requiring a thick Archaean crust creates an apparent dilemma: preserved Archaean lithosphere is characterized by unusually thin crust⁶ (Fig. 3a). A possible solution is crustal delamination, in which crustal thickening leads to a lower crust of dense garnet pyroxenite that can break off and founder into the mantle—triggering an influx of hot asthenosphere that leads to lower-crustal anatexis, crustal uplift and magmatism²³. Overthickening of

mafic lower crust thus leads to delamination, crustal thinning and further deep crustal magmatism. Indeed, previous workers have suggested that lower-crustal delamination was more prevalent in the Archaean, a process that would have been facilitated by high Archaean mantle potential temperatures^{21,24}. Furthermore, because delamination involves a pressure- and temperature-dependent critical crustal thickness²⁴, delamination may relate gradual secular cooling to rapid change in the geochemical record. Archaean delamination thus connects high-degree mantle melting, TTG abundance, dramatic geochemical shifts near the Archaean/Proterozoic boundary and thin preserved Archaean crust.

Although the timing of the observed geochemical transition has an uncertainty of about ± 100 Myr, our results nonetheless reveal a striking temporal coincidence between an important discontinuity in the character of preserved continental lithosphere and the atmospheric great oxidation event (GOE) near the Archaean/Proterozoic boundary (Figs 1–3). It is generally agreed that the atmosphere was mostly anoxic before ~ 2.4 Gyr, as inferred from measurements of mass-independent sulphur isotope fractionation in pyritic shales²⁵, occurrence of banded iron formations²⁶, and the presence of reduced detrital grains (sulphides) in Archaean sediments²⁷ (Fig. 3e). But the cause of subsequent atmospheric oxidation remains a matter of debate; some indications suggest that the GOE did not occur for at least several hundred million years after the evolution of photosynthesis, leading some to argue that an additional geological stimulus was required to trigger widespread atmospheric oxidation^{28,29}.

Proposed mechanisms that could link the deep Earth to atmospheric oxidation at the end of the Archaean include a change in magma degassing style (subaerial versus subaqueous)²⁸, or a change in mantle oxygen fugacity²⁹, both of which would influence the atmosphere via the flux of reduced volcanic volatiles. The former process is possible, but it is not immediately clear how this mechanism would link to our observed geochemical trends. As for the latter process, studies of mantle redox proxies such as V/Sc, including our own analysis (Supplementary Fig. 11), indicate that upper-mantle oxygen fugacity has remained roughly constant over the past 3.8 Gyr (ref. 12). However, despite providing useful constraints on mantle oxygen fugacity, redox tracers such as V/Sc best record changes in oxygen fugacity of the mantle/primitive melt system, not subsequent magma fractionation and evolution; consequently, constraints on the redox state of magmas during evolution and fractionation in the crust are largely absent³⁰.

Magmatic redox fractionation is controlled by the ratio of mineral/melt distribution coefficients, $D^{\text{Fe}^{3+}}/D^{\text{Fe}^{2+}}$, which are very poorly characterized. However, as shown by experimental iron speciation studies, eclogitic minerals such as garnet and sodic clinopyroxene can accommodate substantial Fe^{3+} , and have moderately elevated $\text{Fe}^{3+}/\text{Fe}^{2+}$ ratios compared with other common restitic minerals such as olivine^{31,32}. Consequently, melt extraction from an eclogitic or garnet pyroxenitic restite, as is expected during TTG genesis, could lead to lower melt $\text{Fe}^{3+}/\text{Fe}^{2+}$ relative to melt extracted from a low-pressure restite rich in olivine (which contains substantial Fe but very low $\text{Fe}^{3+}/\text{Fe}^{2+}$)^{31,32} and plagioclase (which contains high $\text{Fe}^{3+}/\text{Fe}^{2+}$ but very little total Fe)³². TTG magmatism may therefore leave more oxidizing power in eclogitic residue in the form of Fe^{3+} , while more reducing power is transported to the surface in the form of Fe^{2+} and equilibrated volatile species such as H_2S , CH_4 and CO .

In this context, Archaean deep crustal magmatism provides a potentially significant influence on surficial oxygen fugacity: simple stoichiometric calculations show that a 15% change in restitic $\text{Fe}^{3+}/\Sigma\text{Fe}$ for a modest magmatic flux of $5 \text{ km}^3 \text{ yr}^{-1}$ would result in an increase in magma reducing capacity by approximately $4.0 \times 10^{12} \text{ mol e}^- \text{ yr}^{-1}$, similar in magnitude to the entire present-day volcanic volatile reducing flux (Methods). Although it is not simple to determine if the termination of this additional reducing flux at the end of the Archaean would be sufficient to trigger atmospheric oxidation, the magnitude of the flux may be constrained by further study of restitic iron partitioning.

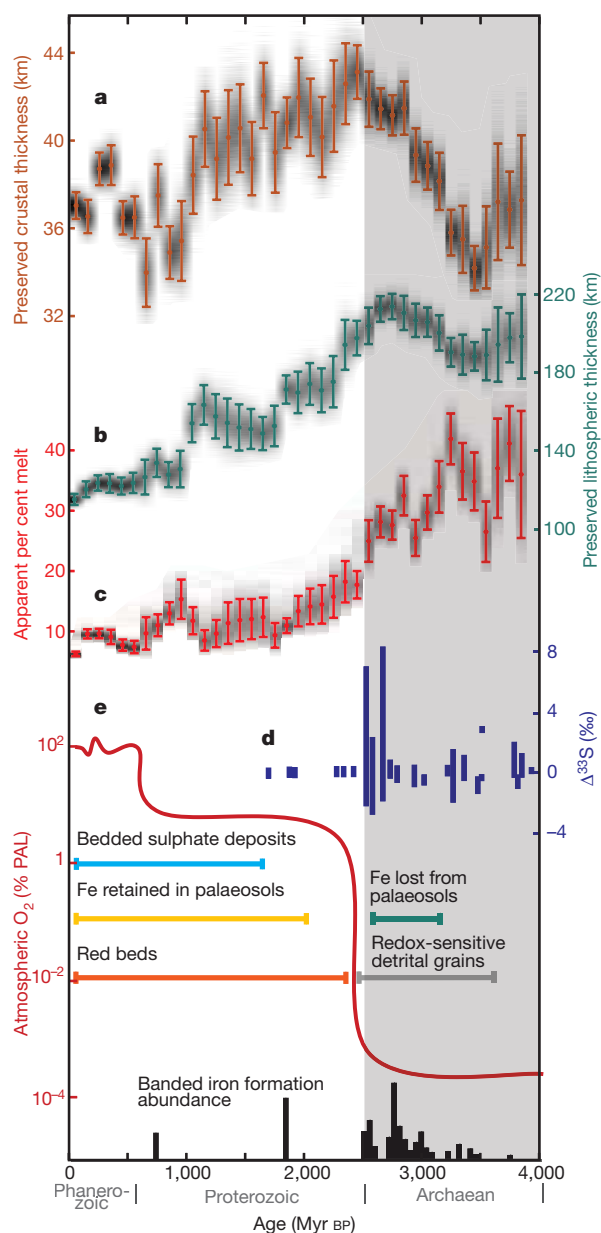


Figure 3 | Relationship between lithospheric evolution and atmospheric oxidation. A notable temporal coincidence is observed among shifts in preserved crustal and lithospheric thickness (this study, based on the models of refs 5 and 14); average degree of mantle melting (this study); and atmospheric oxidation, as shown by a generalized atmospheric oxygen curve (adapted from ref. 7), $\Delta^{33}\text{S}$ fractionation²⁵, and the presence of sedimentary redox indicators^{26,27,33}.

Though speculative, restitic redox partitioning provides a testable mechanism to explain the correlation of atmospheric oxidation with the ~2.5-Gyr geochemical transition revealed by our analysis. It is not yet clear whether these changes represent a fundamental shift in the style of crust formation and tectonics, or simply an abrupt response of melt production to gradual secular cooling. However, delamination and subsequent lower-crustal anatexis provide potential explanations for the observed geochemical trends in felsic rocks that are largely independent of the style of Archaean tectonics. Our analysis illustrates the opportunity presented by recent efforts to make large amounts of published geochemical data readily available and amenable to statistical analysis. These initial results substantiate the existence of a period of profound change in the Earth system near the Archaean/Proterozoic boundary, with implications for crustal differentiation, atmospheric oxidation and the habitability of the Earth for complex aerobic biota.

METHODS SUMMARY

We assembled a database of major- and trace-element analyses of ~70,000 igneous whole-rock samples from a number of pre-existing sources^{3,4,11,13,20}. Each sample was paired with a geospatial location, supplemented by new estimates when necessary. Geochemical data for individual samples were then coupled with crustal and lithospheric thickness, and crustal seismic velocities (v_p and v_s) and density, by projecting sample locations onto high-spatial-resolution global crustal and lithospheric models^{5,14}. Sampling and preservation bias were mitigated by weighted bootstrap resampling, with sample weights inversely proportional to spatiotemporal sample density (see full methods). Estimates of all variables of interest were generated by Monte Carlo analysis as follows: (1) A subset of the data was randomly selected so that the probability of inclusion in the resampled subset is directly proportional to sample weight. (2) A synthetic data set for each sampled data point was drawn from a Gaussian distribution with a mean equal to the original value of the data point and standard deviation equal to the 1σ uncertainty of the data point. (3) The resulting data were sorted into 100-Myr bins, with a mean and variance calculated for each variable. (4) Steps 1–3 were repeated 10,000 times. (5) A total mean and standard error of the mean were calculated for each variable in each bin. Mantle melt fraction (per cent melt) was estimated for each sample in the range 43–51% SiO₂, in all 10,000 simulations, by numerical least-squares optimization of the simulated major-element concentrations to those for slightly hydrous (0.1%) primitive mantle partial melts, as calculated by MELTS^{15,16} for isobaric melting at a range of pressures. The resulting trends in per cent melt estimates agree regardless of melting pressure, suggesting that the assumption of an arbitrary constant pressure does not bias the results.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 November 2011; accepted 5 March 2012.

- Herzberg, C. *et al.* Temperatures in ambient mantle and plumes: constraints from basalts, picrites, and komatiites. *Geochem. Geophys. Geosys.* **8**, 2006GC001390 (2007).
- Hawkesworth, C. J. & Kemp, A. I. S. Evolution of the continental crust. *Nature* **443**, 811–817 (2006).
- Condie, K. C. & O'Neill, C. The Archaean-Proterozoic boundary: 500 my of tectonic transition in Earth history. *Am. J. Sci.* **310**, 775–790 (2010).
- Moyen, J.-F. The composite Archaean grey gneisses: petrological significance, and evidence for a non-unique tectonic setting for Archaean crustal growth. *Lithos* **123**, 21–36 (2011).
- Artemieva, I. M. Global 1 × 1 thermal model TC1 for the continental lithosphere: implications for lithosphere secular evolution. *Tectonophysics* **416**, 245–277 (2006).
- Durrheim, R. J. & Mooney, W. D. Archaean and Proterozoic crustal evolution: evidence from crustal seismology. *Geology* **19**, 606–609 (1991).
- Kump, L. R. The rise of atmospheric oxygen. *Nature* **451**, 277–278 (2008).
- Gando, A. *et al.* Partial radiogenic heat model for Earth revealed by geoneutrino measurements. *Nature Geosci.* **4**, 647–651 (2011).
- Condie, K. C. & Pease, V. (eds) *When Did Plate Tectonics Begin on Planet Earth?* (GSA, 2008).
- de Wit, M. J. On Archaean granites, greenstones, cratons and tectonics: does the evidence demand a verdict? *Precamb. Res.* **91**, 181–226 (1998).
- Condie, K. C. Did the character of subduction change at the end of the Archaean? Constraints from convergent-margin granitoids. *Geology* **36**, 611–614 (2008).
- Li, Z.-X. A. & Lee, C.-T. A. The constancy of upper mantle fO₂ through time inferred from V/Sc ratios in basalts. *Earth Planet. Sci. Lett.* **228**, 483–493 (2004).
- EarthChem. <<http://www.earthchem.org/>> (accessed 14 February 2011).
- Bassin, C., Laske, G. & Masters, G. The current limits of resolution for surface wave tomography in North America. *Eos* **81**, F897 (2000).
- Ghiorso, M. S. & Sack, R. O. Chemical mass transfer in magmatic processes IV. A revised and internally consistent thermodynamic model for the interpolation and extrapolation of liquid-solid equilibria in magmatic systems at elevated temperatures and pressures. *Contrib. Mineral. Petrol.* **119**, 197–212 (1995).
- Asimow, P. D. & Ghiorso, M. S. Algorithmic modifications extending MELTS to calculate subsolidus phase relations. *Am. Mineral.* **83**, 1127–1132 (1998).
- Walter, M. J. Melting of garnet peridotite and the origin of komatiite and depleted lithosphere. *J. Petrol.* **39**, 29–60 (1998).
- Hirschmann, M. M., Ghiorso, M. S., Wasylenski, L. E., Asimow, P. D. & Stolper, E. M. Calculation of peridotite partial melting from thermodynamic models of minerals and melts. I. Review of methods and comparison with experiments. *J. Petrol.* **39**, 1091–1115 (1998).
- Drummond, M. S. & Defant, M. J. A model for trondhjemite-tonalite-dacite genesis and crustal growth via slab melting: Archaean to modern comparisons. *J. Geophys. Res.* **95**, 21,503–21,521 (1990).
- Moyen, J.-F. & Stevens, G. in *Archaean Geodynamics and Environment* (eds Benn, K., Mareschal, J.-C. & Condie, K. C.) 147–175 (AGU Monograph, 2006).
- Petford, N. & Atherton, M. Na-rich partial melts from newly underplated basaltic crust: the Cordillera Blanca Batholith, Peru. *J. Petrol.* **37**, 1491–1521 (1996).
- Müntener, O., Kelemen, P. B. & Grove, T. L. The role of H₂O during crystallization of primitive arc magmas under uppermost mantle conditions and genesis of igneous pyroxenites: an experimental study. *Contrib. Mineral. Petrol.* **141**, 643–658 (2001).
- Kay, R. W. & Mahlborg Kay, S. Delamination and delamination magmatism. *Tectonophysics* **219**, 177–189 (1993).
- Bédard, J. H. A catalytic delamination-driven model for coupled genesis of Archaean crust and sub-continental lithospheric mantle. *Geochim. Cosmochim. Acta* **70**, 1188–1214 (2006).
- Farquhar, J., Bao, H. & Thiemens, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–758 (2000).
- Fischer, W. W. & Knoll, A. H. An iron shuttle for deepwater silica in Late Archaean and early Paleoproterozoic iron formation. *Geol. Soc. Am. Bull.* **121**, 222–235 (2009).
- Rasmussen, B. & Buick, R. Redox state of the Archaean atmosphere: evidence from detrital heavy minerals in ca. 3250–2750 Ma sandstones from the Pilbara Craton, Australia. *Geology* **27**, 115–118 (1999).
- Kump, L. R. & Barley, M. E. Increased subaerial volcanism and the rise of atmospheric oxygen 2.5 billion years ago. *Nature* **448**, 1033–1036 (2007).
- Kasting, J., Egglar, D. & Raeburn, S. Mantle redox evolution and the oxidation state of the Archaean atmosphere. *J. Geol.* **101**, 245–257 (1993).
- Lee, C.-T. A. *et al.* The redox state of arc mantle using Zn/Fe systematics. *Nature* **468**, 681–685 (2010).
- Canil, D. *et al.* Ferric iron in peridotites and mantle oxidation states. *Earth Planet. Sci. Lett.* **123**, 205–220 (1994).
- van Aken, P. V. & Liebscher, B. Quantification of ferrous/ferric ratios in minerals: new evaluation schemes of Fe L 23 electron energy-loss near-edge spectra. *Phys. Chem. Miner.* **29**, 188–200 (2002).
- Rye, R. & Holland, H. D. Paleosols and the evolution of atmospheric oxygen: a critical review. *Am. J. Sci.* **298**, 621–672 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Condie and J.-F. Moyen for providing their data sets of granitoid rocks through time; F. Simons for assistance with statistical methods; C.-T. Lee, W. Fischer, A. Maloof, C. Langmuir, O. Müntener, J. Higgins, A. Rubin, T. Duffy, K. Samperton, and B. Dyer for discussions; and W. White for comments. C.B.K. was supported by a Princeton University Centennial Fellowship.

Author Contributions Both authors interpreted the results and prepared the manuscript. C.B.K. compiled the data set and performed the calculations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.B.K. (cbkeller@princeton.edu).

METHODS

Database compilation. We assembled a database of major- and trace-element analyses of ~70,000 igneous whole-rock samples from a number of pre-existing sources, including ~2,500 from Condie^{3,11}, ~1,500 from Moyen^{4,20}, and the remaining majority from the EarthChem repository (encompassing samples from Navdat, Georoc and the US Geological Survey)¹³. Each sample was paired with a geospatial location, with pre-existing location data supplemented by new estimates when necessary. Geochemical data for individual samples were then coupled with crustal and lithospheric thickness, and crustal v_p , v_s and density, by superimposing sample locations onto high-spatial-resolution global crustal and lithospheric models CRUST 2.0¹⁴ and TC1⁵, respectively. To minimize the chance of introducing any unintentional selection bias, outliers were not removed from the database except in cases where the recorded data were demonstrably unphysical. The result was a data set including up to 98 defined variables (for example, sample age, SiO₂, La, crustal thickness, etc.) for each of the ~70,000 samples, for a total of more than 3.9 million defined data points.

Monte Carlo analysis. Estimates of all variables of interest were generated by Monte Carlo analysis with weighted bootstrap resampling, as illustrated for one case in Supplementary Fig. 1. Sample weights were assigned to be inversely dependent on spatiotemporal sample density, according to the relationship

$$W_i \propto 1 / \sum_{j=1}^n \left(\frac{1}{((z_i - z_j)/a)^2 + 1} + \frac{1}{((t_i - t_j)/b)^2 + 1} \right) \quad (1)$$

where n is the number of samples in the database, z is spatial location, t is age of the rock, and a and b are normalization coefficients of 1.8 arc degrees (200 km) and 38 Myr, respectively. After calculation of weight W_i for each sample i in the database, bootstrap resampling was carried out in the following steps: (1) A subset of data was randomly selected so that the probability of inclusion in the resampled subset is directly proportional to sample weight. (2) A synthetic data set for each sampled data point was drawn from a Gaussian distribution with a mean equal to the original value of the data point and standard deviation equal to the estimated 1σ uncertainty of the data point. (3) The resulting data were sorted into 100-Myr bins, with a mean and variance calculated for each variable. (4) Steps 1–3 were repeated 10,000 times. (5) A total mean and standard error of the mean were calculated for each variable in each bin. To distinguish between processes of mantle melting and subsequent evolution and fractionation, the entire procedure was conducted separately for samples from each of four silica ranges (43–51%, 51–62%, 62–74% and 74–80%).

Owing to the large quantity of random numbers required for Monte Carlo simulation, randomness in steps 1 and 2 was simulated with the Mersenne Twister pseudorandom number generator²⁴, as implemented by MATLAB. To assess the efficacy of the resampling procedure, prior and posterior density distributions were compared to the ideal case for latitude, longitude and time (Supplementary Fig. 2). The results suggest substantial improvement in sampling quality, although tradeoffs exist between uniform sampling through time and uniform sampling in space. To evaluate the applicability of mean and standard error as appropriate statistics to characterize the results of Monte Carlo simulation, the distribution of sample means was examined and found to be approximately Gaussian, as illustrated for one variable in Supplementary Fig. 3. This is expected as a consequence of the Central Limit Theorem; regardless of the distribution of a data set, the distribution of the mean of the data set will naturally tend towards a Gaussian distribution. Finally, to mitigate the undesirable statistical properties of ratios with small denominators and minimize the effect of outliers, geochemical ratios such as La/Yb were calculated first as fractions (for example, La/(La + Yb)) during Monte Carlo analysis and averaging, and subsequently converted to ratios for presentation; the resulting 'fractional' averages are more robust than but not directly comparable to the average of a simple ratio.

In contrast to the widely discussed episodic nature of crust formation (for example, ref. 35), our analysis attempts to produce a uniform record by reducing the weight of abundantly sampled time periods or terranes. For instance, in accordance with equation (1), rock samples from a time period poorly represented in the database will be bootstrap-resampled more heavily than those from periods where much crust is preserved. However, as equation (1) attempts to balance optimal sampling in space and optimal sampling through time, periods where very few data are preserved (for example, 2.6–2.2 Gyr) will remain more poorly sampled, resulting in larger standard errors of the mean at these times. Additionally, spatiotemporal sample weighting in conjunction with Monte Carlo sampling of rocks with large reported age uncertainties result in a continuous geochemical record that smooths out stochastic and stepwise transitions; the slopes of any abrupt trends presented herein are therefore minimum estimates. The nature of the Monte Carlo sampling process has the additional consequence

that there is no uniform or single-valued age uncertainty for the resulting trends. However, temporal uncertainty in the results is generally on the order of 100 Myr.

Response of the Monte Carlo technique to the distribution of the uncertainty of the data. Although our Monte Carlo analysis does not assume any specific statistical distribution for the geochemical, geochronological or geophysical data, it is necessary to specify the distribution of the uncertainties associated with each individual data point in our database. As this uncertainty will be a product of the convolution of multiple independent error sources, it is reasonable to conclude that the distribution of the uncertainty will approach a Gaussian distribution, in accordance with the Central Limit Theorem. (A similar rationale explains why the distribution of the mean of the data approaches a Gaussian distribution (Supplementary Fig. 3), even though the distribution of the data is not Gaussian.) In the sources from which we compiled the database, geochronological data are implied to be Gaussian in accordance with standard practice; no uncertainty is reported for individual geochemical data, but uncertainties for the analytical methods used are typically on the order of a few per cent. Consequently, we have conducted Monte Carlo analysis using the reported geochronological uncertainty for each data point, and assuming a $\pm 2\sigma$ uncertainty of 4% with a Gaussian distribution for the geochemical data, as described in 'Monte Carlo analysis', above. To ensure that this choice of distribution does not bias the distribution of the results, two numerical experiments were carried out.

First, to demonstrate that the choice of error distribution used in Monte Carlo analysis does not bias the distribution of the data or its mean in the one-dimensional case, three Monte Carlo experiments were conducted using: (1) a Gaussian distribution; (2) a log-normal distribution; and (3) a χ^2 distribution with 1 degree of freedom. Using two test data sets (MgO content of mafic igneous samples between 400 and 410 Myr, and a synthetic log-normal distribution), a 10,000-run Monte Carlo simulation was then conducted using each distribution for each data set to simulate the effect of 4% (2σ) analytical error—the same analytical error as is assumed for geochemical data in the Monte Carlo analysis of our 70,000-sample geochemical data set. As seen in Supplementary Fig. 4, for both the MgO data (a–d) and the simulated distribution (e–h), although the choice of error distribution may influence the small-scale smoothness of the resulting data distribution, there is no bias to the overall distribution (as particularly evident in e–h), and no bias to the mean of the distribution (red vertical lines).

To further demonstrate that the choice of statistical distribution to represent analytical error does not bias our Monte Carlo results, we conducted our full Monte Carlo analysis (as above) for three different combinations of error distributions. Supplementary Fig. 5 shows the results as a typical example for MgO in mafic samples (43–51% SiO₂) using: (a) a Gaussian distribution for both geochemical error and geochronological error (as in Figs 1–3); (b) a Gaussian distribution for geochronological error and a log-normal distribution for geochemical error; and (c) a log-normal distribution for both geochemical error and geochronological error. As seen in Supplementary Fig. 5, the resulting trend in MgO through time is nearly identical in each case. We conclude that the choice of error distribution in our Monte Carlo analysis has little to no effect on the resulting trends.

Response of the Monte Carlo results to regions of missing data. Due to the well known paucity of early Proterozoic igneous rocks³⁶, it may be argued that the position and/or shape of the ~2.5-Gyr geochemical transition that we observe may be influenced by a lack of data in this period. In our data set, this lack of data is particularly pronounced between 2.2 and 2.6 Gyr, a period during which the temporal density of geochemical data is only ~5% that at the pronounced data peaks at 2.2 and 2.7 Gyr. Consequently, we have conducted additional tests to examine the sensitivity of the Monte Carlo technique to periods of sparse data. First, a 10,000-point synthetic data set was generated using a log-normal distribution, with an abrupt shift in the mean and variance of the data at 2.5 Gyr. Then, varying amounts of data were deleted in the period from 2.2 to 2.6 Gyr, ranging from 100% of data present to 0% of data present (all deleted). The data set was then subjected to Monte Carlo analysis as above, using an average age uncertainty of 22 Myr and 4% analytical error. As seen in Supplementary Fig. 6, an abrupt transition, as seen in Fig. 1g and Fig. 2b–d, may be preserved to some extent down to the point of 5% data remaining; with 1% and 0% of the data preserved, only a smooth transition is observed. Similarly, deleting data from a continuous starting distribution will not result in any discontinuity in the resulting Monte Carlo output, only a large increase in uncertainty in the data-poor region (Supplementary Fig. 7). Thus, it appears unlikely that the observed abrupt transitions in secular geochemical evolution could be a product of the scarcity of early-Proterozoic data. More gradual transitions such as in Fig. 1f may be influenced by a lack of data; although such trends still provide convincing evidence for a major geochemical transition, the timing and rapidity of that transition is less certain. Although the dearth of data between ~2.2 and ~2.6 Gyr may contribute in cases such as Fig. 1f to the uncertainty in the timing of the observed geochemical transition, it remains the case that between the uncertainty in the timing of this transition and the

uncertainty in the timing of the Great Oxidation Event, these two events easily overlap within uncertainty.

Interpretation of major element trends. Assuming our data set provides a representative sample of the crust, four candidate mechanisms could explain the observed trends (Fig. 1) in compatible and incompatible element composition of mafic rocks: (1) similar degree melting from a mantle that is more enriched now than in the past; (2) higher degrees of crustal contamination of otherwise identical basalts in the present; (3) higher degrees of fractionation in the present; or (4) higher mantle melt fraction in the past, due to either higher temperature or higher water content. A modern-day mantle that is more enriched now than in the Archaean is inconsistent with well established evidence from Nd, Sr and Hf isotope studies of mid-ocean ridge basalts^{37,38}, eliminating mechanism (1). Increasing crustal contamination (2) should be reflected in ratios such as Nb/U and Pb/Ce³⁹, which are instead relatively constant through time (Supplementary Fig. 8). If increasing magma fractionation (3) were used to account for the entire observed trend, this would imply that modern basalts are on average created through fractionation of very-high-MgO magmas such as komatiites, which is clearly not the case. Indeed, if anything, a thick Archaean crust as suggested by crustal depth-of-melting/fractionation indicators (Fig. 2), as well as rapidly cooling high-Mg mafic magmas, suggest increased fractional crystallization in the Archaean, in opposition to the observed trend. To address the issue more quantitatively, we mimic the approach of ref. 40, by applying our bootstrap resampling approach to reconstruct global mean MgO–Na₂O and MgO–CaO/Al₂O₃ trends for time periods between the Archaean and the present day (Supplementary Fig. 9). Both trends are consistent with globally averaged fractionation paths, illustrating that substantial fractionation occurs at any given time, but parallel trends are shifted as a function of age. Following Klein and Langmuir⁴⁰, these parallel trends illustrate changes in mantle melt fraction. Given that decreasing average mantle melt fraction through time (4) is then the most compatible interpretation of the data, we are faced with two options: secular cooling or mantle dehydration. Evidence for a more hydrous Archaean mantle is equivocal⁴¹, and it has even been argued that the potentially undercompensated flux of water into the mantle at subduction zones suggests that the hydration state of the mantle is increasing through time, not decreasing^{42,43}. Secular cooling, on the other hand, presents no such caveats and is widely expected on the basis of physical constraints.

The degree of mantle melting was estimated for each sample with 43–51% SiO₂, in all 10,000 simulations, by numerical least squares optimization of the simulated major-element data to that of slightly hydrous (0.1%) primitive mantle partial melts as calculated by MELTS^{15,16} for isobaric melting at a range of pressure conditions. Resulting per cent melt estimates were found to agree within error despite order of magnitude differences in melting pressure (Supplementary Fig. 10), suggesting that the results are insensitive to changes in model pressure. Although MELTS does not extend to high enough pressures to directly simulate high-degree isentropic decompression melting, as long as the system is not far from equilibrium, per cent melt and melt composition will be roughly state functions of pressure and temperature—that is, path-independent. Consequently, at a given pressure (say, corresponding to the base of the crust) per cent melt and melt composition will both be state functions of temperature. In this case, then, there will be a simple relationship between per cent melt and melt composition, allowing for accurate estimation of per cent melt from melt composition regardless of the exact melting path. This approach will break down in the case of fractional melt extraction, in which the melt cannot be in equilibrium with the solid; the absolute value of the results may therefore be inaccurate in the case of predominantly fractional melting, though relative trends would be expected to remain robust.

Because our analysis focuses on the trends in melt percentage over Earth history rather than the absolute values of these estimates, a number of sources of systematic error can be disregarded. We have intentionally not corrected basalt major-element compositions for fractional crystallization to calculate primary melt compositions, as such an analysis has already been conducted on smaller data sets, and is only possible for a small proportion of the available data⁴⁴. By neglecting the effect of fractional crystallization, our estimates of per cent melting through time are systematic underestimates. Nonetheless, in the absence of large changes in the degree of fractional crystallization through time, as is argued above, the results should accurately represent changes in per cent melt over time. Herzberg *et al.*⁴⁴ analysed a database of ~1,500 basalts through early Earth history, and identified those that could be corrected to primary mantle melts ($n = 33$) to extract mantle potential temperature. The topology of a temperature curve corresponding to our degree-melting curve (Fig. 1f) agrees well with Herzberg *et al.*'s mantle potential temperature, although, as expected, they obtain higher mantle temperatures⁴⁴. Moreover, our per cent melting results generally agree with independent estimates of modern

and Archaean percent melting, suggesting that the degree of underestimation is not large^{17,45}.

To determine if observed trends in trace-element ratios such as La/Yb in low-silica samples could be generated solely by a single set of bulk partition coefficients given the estimated per cent melting curve, we used linearized least-squares regression to optimize bulk partition coefficients to fit per cent melting and trace-element curves with analytical non-modal batch melting equations. No single set of partition coefficients can reproduce the observed trace-element ratios in basalts by varying melt fraction alone, showing that other variables such as temperature and mineralogy may have influenced partition coefficients, or that significant changes in source trace-element composition (for example, for the example of an arc setting, change in the trace-element composition of the mantle wedge due to subduction erosion or sediment subduction) had an effect.

Redox partitioning calculations. To assess the potential for magmatic redox partitioning to affect surficial redox state, a simple stoichiometric redox balance of mantle restite was calculated for a modest magmatic flux of 5 km³ yr⁻¹ (equivalent to estimates of modern arc magmatism⁴⁶), assuming a magma density of 2,500 kg m⁻³, 6 wt % restitic Fe, and a ratio of 2 kg of restite to 1 kg of magma, as follows:

$$\frac{5 \text{ km}^3 \text{ magma}}{\text{yr}} \times \frac{1,000^3 \text{ m}^3}{\text{km}^3} \times \frac{2,500 \text{ kg}}{\text{m}^3} \times \frac{0.06 \text{ kg Fe}}{\text{kg magma}} \times \frac{2 \text{ kg restite}}{\text{kg magma}} \times \frac{1,000 \text{ g}}{\text{kg}} \times \frac{1 \text{ mol Fe}}{55.85 \text{ g Fe}} = \frac{2.68 \times 10^{13} \text{ mol Fe}}{\text{yr}} \quad (2)$$

Consequently, a 15% increase in Fe³⁺/ΣFe of this restitic flux would liberate 4.02 × 10¹² mol e⁻ yr⁻¹, or, stoichiometrically, consume 1.0 × 10¹² mol O₂ yr⁻¹—similar in magnitude to estimates of the entire present-day volcanic volatile flux of ~5.6 × 10¹² mol e⁻ yr⁻¹ (ref. 46).

Notably, the Fe³⁺/Fe²⁺ ratio of many magmas is low even today, apparently not leaving much room for variation through time. However, theory and experiment suggest that oxygen fugacity is proportional to approximately the fourth power of ferric/ferrous ratio⁴⁷. Thus, a change in ferric/ferrous ratio from 0.1 to 0.01 would result in a decrease of roughly ten orders of magnitude in magmatic oxygen fugacity, with potentially important consequences for the production of volcanic volatiles.

34. Matsumoto, M. & Nishimura, T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.* **8**, 3–30 (1998).
35. Condie, K. C., Bickford, M. E., Aster, R. C., Belousova, E. & Scholl, D. W. Episodic zircon ages, Hf isotopic composition, and the preservation rate of continental crust. *Geol. Soc. Am. Bull.* **123**, 951–957 (2011).
36. Condie, K. C. & Aster, R. C. Episodic zircon age spectra of orogenic granitoids: The supercontinent connection and continental growth. *Precamb. Res.* **180**, 227–236 (2010).
37. White, W. M. & Hofmann, A. W. Sr and Nd isotope geochemistry of oceanic basalts and mantle evolution. *Nature* **296**, 821–825 (1982).
38. Vervoort, J. D. & Blichert-Toft, J. Evolution of the depleted mantle: Hf isotope evidence from juvenile rocks through time. *Geochim. Cosmochim. Acta* **63**, 533–556 (1999).
39. Hofmann, A. W., Jochum, K. P., Seufert, M. & White, W. M. Nb and Pb in oceanic basalts: new constraints on mantle evolution. *Earth Planet. Sci. Lett.* **79**, 33–45 (1986).
40. Klein, E. & Langmuir, C. Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
41. Berry, A. J., Danyushevsky, L. V., O'Neill, H. S. C., Newville, M. & Sutton, S. R. Oxidation state of iron in komatiitic melt inclusions indicates hot Archaean mantle. *Nature* **455**, 960–963 (2008).
42. Peacock, S. M. Fluid processes in subduction zones. *Science* **248**, 329–337 (1990).
43. Maruyama, S. & Okamoto, K. Water transportation from the subducting slab into the mantle transition zone. *Gondwana Res.* **11**, 148–165 (2007).
44. Herzberg, C., Condie, K. & Korenaga, J. Thermal history of the Earth and its petrological expression. *Earth Planet. Sci. Lett.* **292**, 79–88 (2010).
45. Walter, M. J. Melt extraction and compositional variability in mantle lithosphere. in *Treatise on Geochemistry*. Vol. 1 (eds Holland, H. D. & Turekian, K. K.) 363–394 (Pergamon, 2003).
46. Kump, L., Kasting, J. & Barley, M. Rise of atmospheric oxygen and the “upside-down” Archaean mantle. *Geochim. Geophys. Geosys.* **2**, 2000GC000114 (2001).
47. Nikolaev, G. S., Borisov, A. A. & Ariskin, A. A. Calculation of the ferric-ferrous ratio in magmatic melts: testing and additional calibration of empirical equations for various magmatic series. *Geochim. Int.* **34**, 641–649 (1996).
48. Crisp, J. A. Rates of magma emplacement and volcanic output. *J. Volcanol. Geotherm. Res.* **20**, 177–211 (1984).

Warming experiments underpredict plant phenological responses to climate change

E. M. Wolkovich¹, B. I. Cook^{2,3}, J. M. Allen⁴, T. M. Crimmins⁵, J. L. Betancourt⁶, S. E. Travers⁷, S. Pau⁸, J. Regetz⁸, T. J. Davies⁹, N. J. B. Kraft^{10,11}, T. R. Ault¹², K. Bolmgren^{13,14}, S. J. Mazer¹⁵, G. J. McCabe¹⁶, B. J. McGill¹⁷, C. Parmesan^{18,19}, N. Salamin^{20,21}, M. D. Schwartz²² & E. E. Cleland¹

Warming experiments are increasingly relied on to estimate plant responses to global climate change^{1,2}. For experiments to provide meaningful predictions of future responses, they should reflect the empirical record of responses to temperature variability and recent warming, including advances in the timing of flowering and leafing^{3–5}. We compared phenology (the timing of recurring life history events) in observational studies and warming experiments spanning four continents and 1,634 plant species using a common measure of temperature sensitivity (change in days per degree Celsius). We show that warming experiments underpredict advances in the timing of flowering and leafing by 8.5-fold and 4.0-fold, respectively, compared with long-term observations. For species that were common to both study types, the experimental results did not match the observational data in sign or magnitude. The observational data also showed that species that flower earliest in the spring have the highest temperature sensitivities, but this trend was not reflected in the experimental data. These significant mismatches seem to be unrelated to the study length or to the degree of manipulated warming in experiments. The discrepancy between experiments and observations, however, could arise from complex interactions among multiple drivers in the observational data, or it could arise from remediable artefacts in the experiments that result in lower irradiance and drier soils, thus dampening the phenological responses to manipulated warming. Our results introduce uncertainty into ecosystem models that are informed solely by experiments and suggest that responses to climate change that are predicted using such models should be re-evaluated.

Predicting species' responses to climate change is a major challenge in ecology. Plants have been a focus of study because their responses can affect entire food webs, disturbance regimes and crucial ecosystem services, including pollination, carbon and nutrient cycling, and water supply⁶. Researchers have adopted two main approaches to predict plant responses to climate change. The first approach relies on observations over time or space to quantify sensitivities to temperature variability and change, including recent warming that is at least partly associated with greenhouse gas emissions^{3–5}. The second approach relies on experiments that directly warm natural plant communities on a small scale^{2,7}. Results from both methods suggest that advancing phenology (for example, earlier flowering and leafing) is one of the most sensitive plant responses to warming. Current estimates of changes in phenology are 1.9–3.3 days per °C for experiments¹ and 2.5–5 days per °C for observations^{3,8}.

Warming experiments have been used to extrapolate to future climate conditions for more than 20 years^{7,9}. This approach rests on the critical but little-tested assumption that plant responses to experimental warming match long-term responses to global warming. Testing this assumption is an important first step in assessing the utility of warming experiments for long-term forecasting and prediction. Recent studies have compared experimental and observational results at single sites^{2,10}, and one study found coherence across methods in plant responses to warming¹⁰; however, cross-site comparisons have proved more challenging.

Here we present two new spatially and temporally extensive databases of plant phenology for 1,634 species based on long-term observations and short-term warming experiments (Fig. 1). The databases were developed specifically to test how accurately short-term warming experiments predict the long-term responses of wild plants. We tested for differences in the overall estimated phenological sensitivity to temperature (change in flowering or leafing date per °C, hereafter referred to as 'temperature sensitivity') between the experiments and the observations. The sensitivities for each species at a site, i , were calculated as the β coefficient from the following: phenological event date _{i} = $\beta_i(\text{MAT}) + \varepsilon_i$, where MAT is the mean annual temperature for each site and ε is error. Our observational sensitivities are thus based on interannual temperature variability. For the sensitivities from experiments, we used a regression approach similar to the observational calculation, which simplifies to the following: (phenological event date _{i ,warm} – phenological event date _{i ,control}) / ΔT , where ΔT is the change in the temperature reported between the warmed and the control plots (Supplementary Information). We looked at how the sensitivities to temperature varied across space and time, and we tested whether sensitivity increases with latitude⁵ or with the magnitude of temperature change⁸. We also assessed whether species that flower relatively early in the spring are more sensitive to temperature than later-flowering species, as has been found in several observational³ and experimental^{10,11} studies. Although other abiotic factors, including photoperiod¹² and snowmelt¹³, may affect phenology, we focus on temperature here because it is the most consistent and dominant controller of spring phenology^{12,14}, as well as the most reported.

Warming experiments underpredicted observational responses to climate change (Fig. 2). Across all species, the experiments underpredicted the magnitude of the advance—for both leafing and flowering—that results from temperature increases (Fig. 2a; flowering, $F_{1,34} = 9.7$,

¹Division of Biological Sciences, University of California San Diego, 9500 Gilman Drive 0116, La Jolla, California 92093, USA. ²NASA Goddard Institute for Space Studies, New York, New York 10025, USA. ³Ocean and Climate Physics, Lamont-Doherty Earth Observatory, Palisades, New York 10964-8000, USA. ⁴Department of Ecology & Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269-3043, USA. ⁵USA National Phenology Network, 1955 East Sixth Street, Tucson, Arizona 85721, USA. ⁶US Geological Survey, 1955 East Sixth Street, Tucson, Arizona 85719, USA. ⁷Department of Biological Sciences, North Dakota State University, Fargo, North Dakota 58108, USA. ⁸National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, California 93101, USA. ⁹Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montreal, Quebec H3A 1B1, Canada. ¹⁰Biodiversity Research Centre, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada. ¹¹Department of Biology, University of Maryland, College Park, Maryland 20742, USA. ¹²National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307, USA. ¹³Swedish University of Agricultural Sciences, Swedish National Phenology Network, Asa, Unit for Field-based Forest Research, SE-36030 Lammhult, Sweden. ¹⁴Theoretical Population Ecology and Evolution, Lund University, SE-22362 Lund, Sweden. ¹⁵Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, California 93106, USA. ¹⁶US Geological Survey, Denver Federal Center, MS 412, Denver, Colorado 80225, USA. ¹⁷School of Biology and Ecology & Sustainability Solutions Initiative, University of Maine, Orono, Maine 04469, USA. ¹⁸Integrative Biology, University of Texas, 1 University Station C0930, Austin, Texas 78712, USA. ¹⁹Marine Sciences Institute, A425 Portland Square, Drake Circus, University of Plymouth, Plymouth, Devon PL4 8AA, UK. ²⁰Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland. ²¹Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland. ²²Department of Geography, Bolton 410, PO Box 413, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201-0413, USA.

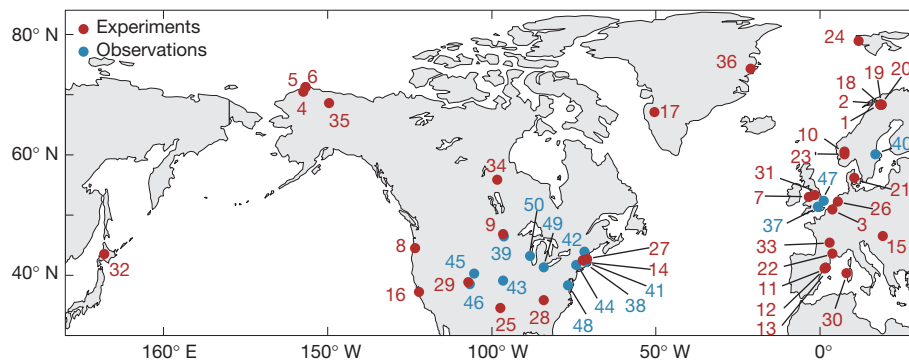


Figure 1 | Map of the experimental and observational sites. The map covers 14 long-term observational phenology studies and 36 experimental phenology studies (one experimental study site, in Australia, is not shown). The numbers correspond to site information given in the Supplementary Information.

$P = 0.004$; leafing, $F_{1,18} = 2.8$, $P = 0.1$). The estimates for observations were based on plant responses to interannual temperature variability, yet they were similar to estimates calculated from long-term trends (Fig. 3) and were consistent across alternative approaches to calculating mean sensitivities (Supplementary Information). Because there are far more species represented in our observational data (1,558 unique taxa) than in the experimental data (115), this difference in mean temperature sensitivities could be attributable to a sampling effect, in which the increased sampling of species adds species that are more sensitive. However, when analyses were restricted to the 36 species common to both study types, the experiments still produced smaller sensitivities (in the sign tests for flowering and leafing, $P = 0.02$, $n = 30$ for flowering and $n = 7$ for leafing). Most strikingly, the experiments predicted a delay in flowering and little change in leafing in response to higher temperatures, whereas the observations predicted an advance of >4.6 days per $^{\circ}\text{C}$ for either phenological event (Fig. 2b).

The experiments also failed to match the observational studies' predictions of greater advances in the flowering of early-season species with warming (Fig. 4; mean flowering date, $F_{1,2114} = 93.2$, $P < 0.0001$; flowering date \times study type, $F_{1,2114} = 3.7$, $P = 0.05$). Across experimental and observational study types, sensitivities were not associated with latitude ($F_{1,31} = 0.02$, $P > 0.8$) or interannual climate variability ($F_{1,31} = 0.1$, $P > 0.7$), suggesting that, although the interspecific variation in sensitivities is great, communities that are governed by different climate regimes do not differ in their overall sensitivity to warming. Additionally, we found no evidence that sensitivities varied with the degree of experimental warming ($F_{1,76} = 1.1$, $P = 0.3$); this is in contrast to the observational studies, which show an increase in sensitivities with

the degree of warming (Supplementary Information; $F_{1,4856} = 12.3$, $P = 0.0005$; see also ref. 8).

Several artefacts might explain the lower sensitivities in the experimental studies, including differences in the traits of the species sampled, correlations between MAT and other environmental variables for the observational studies, and differences in the degree of warming. We found no evidence, however, that these factors explain the smaller or positive sensitivities found in the experiments. Both study types mainly examined longer-lived plant species (85.0% and 88.4% were perennials in the experiments and the observations, respectively), and most were herbs or forbs (78.0% and 72.3%, respectively) not woody species (Supplementary Information). In addition, we found no correlations between MAT and other climate variables for which data were

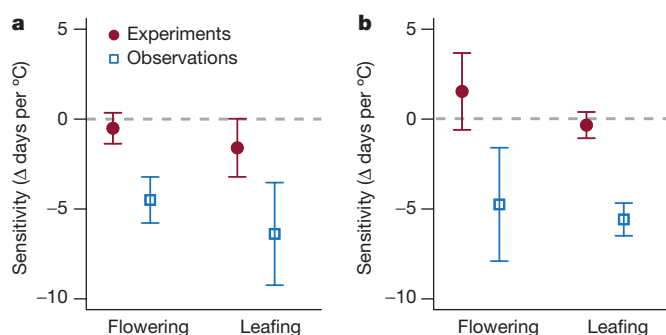


Figure 2 | Estimates of the flowering and leafing sensitivities. The estimates from the mixed effects model (presented as mean \pm s.e.m.), including the random effects of site and species, show that experiments underpredict the magnitude of plant responses to interannual temperature variation for all species sampled (a) and for the species that are common to both the experimental and the observational data sets (b). The region above the dashed grey line represents positive sensitivities, meaning that the species' phenological events are delayed with warming, whereas the region below the line represents negative sensitivities, meaning that the species' events advance with warming.

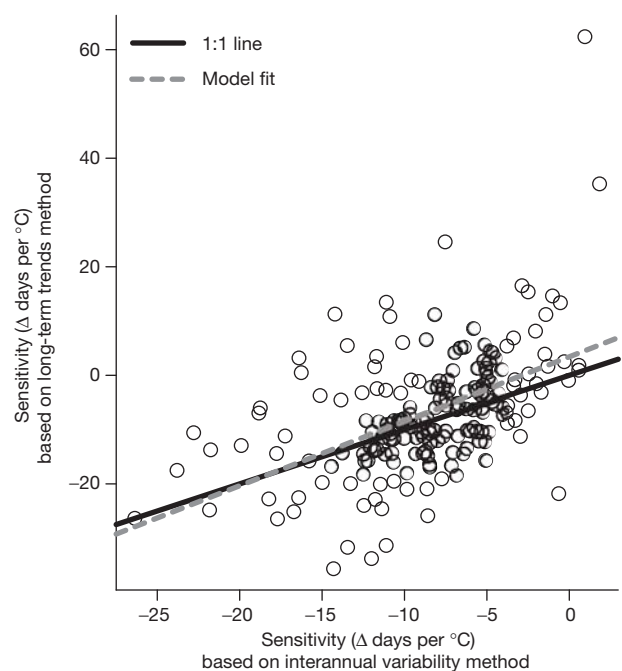


Figure 3 | Relationship between sensitivities to temperature calculated from interannual variation and from long-term trends. Observational sensitivities calculated from interannual variation (regression of flowering date versus MAT) were strongly correlated with those calculated from long-term trends (long-term trend in the flowering date versus long-term trend in MAT). The mixed effects model fit ($F_{1,185} = 63.9$, $P < 0.0001$) and the 1:1 line are both shown. For long-term sensitivity estimates, we used the subset of the NECTAR data corresponding to species with observations between 1971 and 2000 (a common standard for long-term climate trend analysis). See the Supplementary Information (Observational sensitivities: comparisons with long-term trends and 1971–2000 standard) for complete details and for further alternative methods and data for calculating observational sensitivities.

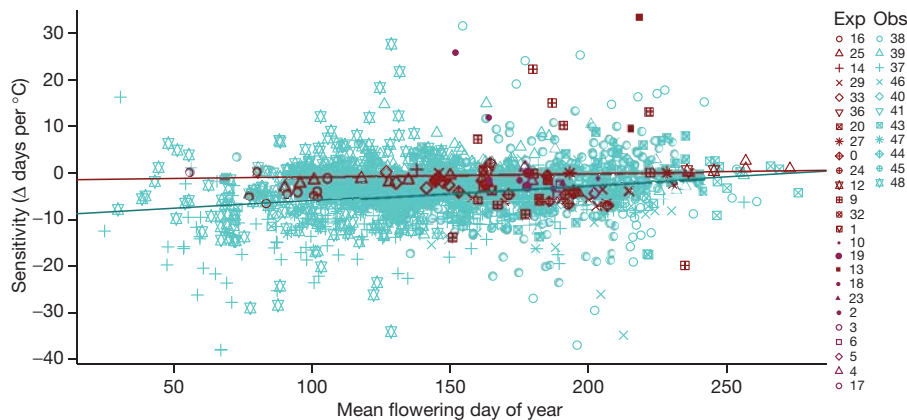


Figure 4 | Sensitivity to temperature decreases throughout the growing season. This trend is seen in the observational studies (blue) but not in the experimental studies (red). The numbers correspond to those in Fig. 1 and to site information given in the Supplementary Information. The linear model

fitted to the observational data predicts that species blooming 3 months apart would have differences in sensitivities of >3 days per $^{\circ}\text{C}$. Exp, experimental data; Obs, observational data.

available (Supplementary Information). Finally, both study types spanned roughly a $0\text{--}5^{\circ}\text{C}$ temperature range, with a median of $1.4\text{--}2.4^{\circ}\text{C}$ (Supplementary Information).

Given the difference in time frames between the experimental and observational studies (3.8 years and 31.0 years, respectively; $F_{1,49} = 72.0$, $P < 0.0001$), the experiments may capture predominantly plastic responses to temperature change, whereas the observational studies may also integrate shifts in genotypes, community composition and ecosystem dynamics. Furthermore, plant responses may have shifted fundamentally following the significant Northern Hemisphere warming in all seasons starting in the 1970s¹⁵, thus the experiments (which have been conducted relatively recently) may capture only these shifted dynamics. However, we found that these scaling issues would create a bias towards detecting greater sensitivities from short-term, post-1970 studies (Supplementary Information), which is opposite to the trends in the experimental studies. Sensitivities that captured 31–35 years of phenological change were generally smaller than those that spanned 5 or fewer years (sign tests of $P \leq 0.02$, Supplementary Information). In addition, the sensitivities were similar before 1970 and after 1970 ($P > 0.1$, Supplementary Information), when warming trends began to accelerate¹⁵.

Although warming methods have continually improved^{16–18}, all experimental manipulations unavoidably alter additional environmental factors. For example, the most common passive warming structures—open-top chambers—reduce light, wind and often soil moisture and can unintentionally increase minimum winter temperatures^{19,20}. By contrast, above-canopy heaters, which are common structures for active warming, often achieve warming only in periods of low wind¹⁷. We found no evidence of differences in the sensitivities between the passive and active designs ($F_{1,23} = 0.05$, $P > 0.8$). Designs using above-canopy heating, however, tended to produce the greatest phenological advances for flowering (Supplementary Information; $F_{2,22} = 7.2$, $P = 0.004$), and this was the type of heating used in the one study that found correspondence between experimental and observational methods¹⁰ (Supplementary Information). There were no effects of habitat (for example, alpine or arctic), study duration or period of warming each year (Supplementary Information): the results from studies that manipulated temperature only during the growing season were indistinguishable from those that elevated temperature year round ($F_{1,18} = 0.5$, $P > 0.4$). Both active and passive experimental designs tend to reduce total irradiance and soil moisture^{11,19,21}, which can delay plant phenology^{2,22}. Thus, the lower sensitivities of the experiments could be the product of planned temperature increases that tend to advance phenology and artefacts that tend to delay phenology. Alternatively, the lower sensitivities of the experiments could occur if experimental studies

isolate temperature effects, whereas observational data integrate the effects of complex and possibly reinforcing interactions among multiple drivers of anthropogenic climate change and variability (Supplementary Information).

One needed improvement for both experimental and observational studies is the increased reporting of temperature and other relevant environmental metrics. Many experimental studies report coarse estimates of temperature differences (for example, annual means), often omitting measurement duration, frequency and statistics. Furthermore, the few efforts to examine temperature changes carefully have found that warming chambers can sometimes result in the cooling of plots²⁰ or can have unintended effects on daily and seasonal temperature regimes¹⁹. Such variation might confound experimental study data because plants are differentially sensitive to temperatures during different parts of the day and integrate temperature differentially over months and seasons²³. Moving towards more relevant measures of temperature sensitivity should also be a future goal for observational studies. For example, using MAT, we found that 23.4% of the species in the observational data responded significantly to temperature, but this proportion nearly tripled (to 67.6%) when a growing-degree day model was used (Supplementary Information). Additionally, researchers collecting experimental or observational data should report other environmental metrics that may drive phenology. For example, spring phenology is driven by snowmelt date in many alpine and tundra habitats^{10,24}. This date is not reported, however, at most sites, and experiments often remove warming structures in the snowpack season, thus not altering snowmelt date. Finally, we found high interspecific variation in sensitivities in both study types, suggesting that increased species sampling is needed for accurate climate change forecasting.

Our estimates of wild plant sensitivity to temperature, based on observations for 1,558 species, suggest that spring leafing and flowering will continue to advance at the rate of 5–6 days per $^{\circ}\text{C}$ in many regions of the world (within the limits of historical temperature ranges)²⁵. The database underlying our estimates represents a sampling of species that is almost an order of magnitude higher than the most recent comprehensive meta-analysis of plant phenology³; yet sensitivity estimates using data derived from ref. 3 are highly similar to those presented here (Supplementary Information). Comparisons with temporal estimates (for example, the change in days per decade^{5,26}) are difficult; however, given a global increase in temperature of 0.6°C over the past four decades¹⁵, our estimates seem to be in line with the current estimates of -1.1 to -3.3 days per decade for plants⁵.

Accurate models of biotic responses to climate change require comparable methods to estimate how plants shift with increasing temperatures. Field-based warming studies allow us to mechanistically test single versus multiple factors and to project forward to novel global

conditions for which historical observations offer no comparisons²⁷. To project experimental results, however, we must first verify that they match the changes that are observed in natural systems. Providing a first step, we have shown that warming experiments underpredict the advance of spring events observed over recent decades. Furthermore, when sampling the same species, the experiments failed to predict both the magnitude and the direction of plant responses to warming. Such differences between observed and experimentally estimated temperature sensitivities indicate that experimental results alone should not be used for parameterizing species distribution and ecosystem models. Although long-term observational data are currently the best measures of biotic responses to anthropogenic climate change²⁸, our results suggest that researchers may not understand these responses well enough to replicate them experimentally. Obvious remedies include ongoing efforts to improve experimental designs and to expand observational networks, improved prediction of species-level variation, and refinement and standardization of temperature sensitivities across experimental and observational studies.

METHODS SUMMARY

Complete methods are given in the Data and Methods sections of the Supplementary Information.

Received 21 November 2011; accepted 6 March 2012.

Published online 2 May 2012.

- Arft, A. M. *et al.* Responses of tundra plants to experimental warming: meta-analysis of the international tundra experiment. *Ecol. Monogr.* **69**, 491–511 (1999).
- Morin, X., Roy, J., Sonie, L. & Chuine, I. Changes in leaf phenology of three European oak species in response to experimental climate change. *New Phytol.* **186**, 900–910 (2010).
- Menzel, A. *et al.* European phenological response to climate change matches the warming pattern. *Glob. Change Biol.* **12**, 1969–1976 (2006).
- Schwartz, M. D., Ahas, R. & Aasa, A. Onset of spring starting earlier across the Northern Hemisphere. *Glob. Change Biol.* **12**, 343–351 (2006).
- Parmesan, C. Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Glob. Change Biol.* **13**, 1860–1872 (2007).
- Cleland, E. E., Chuine, I., Menzel, A., Mooney, H. A. & Schwartz, M. D. Shifting plant phenology in response to global change. *Trends Ecol. Evol.* **22**, 357–365 (2007).
- Chapin, F. S. & Shaver, G. R. Physiological and growth responses of arctic plants to a field experiment simulating climatic change. *Ecology* **77**, 822–840 (1996).
- Amano, T., Smithers, R. J., Sparks, T. H. & Sutherland, W. J. A 250-year index of first flowering dates and its response to temperature changes. *Proc. R. Soc. Lond. B* **277**, 2451–2457 (2010).
- Harte, J. & Shaw, R. Shifting dominance within a montane vegetation community: results of a climate-warming experiment. *Science* **267**, 876–880 (1995).
- Dunne, J. A., Harte, J. & Taylor, K. J. Subalpine meadow flowering phenology responses to climate change: integrating experimental and gradient methods. *Ecol. Monogr.* **73**, 69–86 (2003).
- Sherry, R. A. *et al.* Divergence of reproductive phenology under climate warming. *Proc. Natl Acad. Sci. USA* **104**, 198–202 (2007).
- Chuine, I., Morin, X. & Bugmann, H. Warming, photoperiods, and tree phenology. *Science* **329**, 277–278 (2010).
- Dunne, J. A., Saleska, S. R., Fischer, M. L. & Harte, J. Integrating experimental and gradient methods in ecological climate change research. *Ecology* **85**, 904–916 (2004).
- Rohde, A., Bastien, C. & Boerjan, W. Temperature signals contribute to the timing of photoperiodic growth cessation and bud set in poplar. *Tree Physiol.* **31**, 472–482 (2011).
- Trenberth, K. E. & Josey, S. A. *Observations: Surface and Atmospheric Climate Change* 235–336 (IPCC, 2007).
- Marion, G. M. *et al.* Open-top designs for manipulating field temperature in high-latitude ecosystems. *Glob. Change Biol.* **3**, 20–32 (1997).
- Kimball, B. A. Theory and performance of an infrared heater for ecosystem warming. *Glob. Change Biol.* **11**, 2041–2056 (2005).
- Kimball, B. A. *et al.* Infrared heater arrays for warming ecosystem field plots. *Glob. Change Biol.* **142**, 309–320 (2008).
- Kennedy, A. D. Simulated climate-change: are passive greenhouses a valid microcosm for testing the biological effects of environmental perturbations? *Glob. Change Biol.* **1**, 29–42 (1995).
- Molau, U. Responses to natural climatic variation and experimental warming in two tundra plant species with contrasting life forms: *Cassiope tetragona* and *Ranunculus nivalis*. *Glob. Change Biol.* **3**, 97–107 (1997).
- Kennedy, A. D. Temperature effects of passive greenhouse apparatus in high-latitude climate-change experiments. *Funct. Ecol.* **9**, 340–350 (1995).
- Rathcke, B. & Lacey, E. P. Phenological patterns of terrestrial plants. *Annu. Rev. Ecol. Syst.* **16**, 179–214 (1985).
- Aikawa, S., Kobayashi, M. J., Satake, A., Shimizu, K. K. & Kudoh, H. Robust control of the seasonal expression of the *Arabidopsis* FLC gene in a fluctuating environment. *Proc. Natl Acad. Sci. USA* **107**, 11632–11637 (2010).
- Chapin, F. S., Matson, P. A. & Mooney, H. A. *Principles of Terrestrial Ecosystem Ecology* (Springer, 2002).
- Körner, C. & Basler, D. Phenology under global warming. *Science* **327**, 1461–1462 (2010).
- Root, T. L. *et al.* Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57–60 (2003).
- Williams, J. W. & Jackson, S. T. Novel climates, no-analog communities, and ecological surprises. *Front. Ecol. Environ.* **59**, 475–482 (2007).
- Rosenzweig, C. *et al.* Attributing physical and biological impacts to anthropogenic climate change. *Nature* **453**, 353–357 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was conducted as part of the Forecasting Phenology Working Group supported by the National Center for Ecological Analysis & Synthesis (EF-0553768), with additional support from National Science Foundation grants DBI-0905806, IOS-0639794, DEB-0922080 and the Natural Sciences and Engineering Research Council of Canada CREATE Training Program. Special thanks to the many data managers, including G. Aldridge, P. Huth, D. Inouye, G. Johansson, A. Miller-Rushing, J. O'Keefe, R. Primack, S. Smiley, T. Sparks and J. Thompson. We thank M. Ayres, L. Kueppers, D. Moore and M. O'Connor for comments on earlier drafts.

Author Contributions E.M.W. conceived the idea, performed analyses and wrote the paper. B.I.C. performed analyses. E.E.C. and N.J.B.K. assisted with analyses. E.M.W., E.E.C., J.M.A., T.M.C., S.E.T. and S.P. developed the STONE database. E.M.W., B.I.C. and J.R. contributed extensively to development of the NECTAR database. All authors (including J.L.B., T.J.D., T.R.A., K.B., S.J.M., G.J.M., B.J.M., C.P., N.S. and M.D.S.) contributed to the editing of the manuscript and to data management of the observational data sets.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.M.W. (wolkovich@biodiversity.ubc.ca).

Melanoma genome sequencing reveals frequent *PREX2* mutations

Michael F. Berger^{1†*}, Eran Hodis^{1*}, Timothy P. Heffernan^{2†*}, Yonathan Lissanu Deribe^{2†*}, Michael S. Lawrence¹, Alexei Protopopov^{2†}, Elena Ivanova², Ian R. Watson^{2†}, Elizabeth Nickerson¹, Papia Ghosh², Hailei Zhang², Rhamy Zeid², Xiaojia Ren², Kristian Cibulskis¹, Andrey Y. Sivachenko¹, Nikhil Wagle^{2,3}, Antje Sucker⁴, Carrie Sougnez¹, Robert Onofrio¹, Lauren Ambrogio¹, Daniel Auclair¹, Timothy Fennell¹, Scott L. Carter¹, Yotam Drier⁵, Petar Stojanov¹, Meredith A. Singer^{2†}, Douglas Voet¹, Rui Jing¹, Gordon Saksena¹, Jordi Barretina¹, Alex H. Ramos^{1,3}, Trevor J. Pugh^{1,2,3}, Nicolas Stransky¹, Melissa Parkin¹, Wendy Winckler¹, Scott Mahan¹, Kristin Ardlie¹, Jennifer Baldwin¹, Jennifer Wargo⁶, Dirk Schadendorf⁴, Matthew Meyerson^{1,2,3,7}, Stacey B. Gabriel¹, Todd R. Golub^{1,7,8,9}, Stephan N. Wagner¹⁰, Eric S. Lander^{1,11*}, Gad Getz^{1*}, Lynda Chin^{1,2,3†*} & Levi A. Garraway^{1,2,3,7*}

Melanoma is notable for its metastatic propensity, lethality in the advanced setting and association with ultraviolet exposure early in life¹. To obtain a comprehensive genomic view of melanoma in humans, we sequenced the genomes of 25 metastatic melanomas and matched germline DNA. A wide range of point mutation rates was observed: lowest in melanomas whose primaries arose on non-ultraviolet-exposed hairless skin of the extremities (3 and 14 per megabase (Mb) of genome), intermediate in those originating from hair-bearing skin of the trunk (5–55 per Mb), and highest in a patient with a documented history of chronic sun exposure (111 per Mb). Analysis of whole-genome sequence data identified *PREX2* (phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchange factor 2)—a PTEN-interacting protein and negative regulator of PTEN in breast cancer²—as a significantly mutated gene with a mutation frequency of approximately 14% in an independent extension cohort of 107 human melanomas. *PREX2* mutations are biologically relevant, as ectopic expression of mutant *PREX2* accelerated tumour formation of immortalized human melanocytes *in vivo*. Thus, whole-genome sequencing of human melanoma tumours revealed genomic evidence of ultraviolet pathogenesis and discovered a new recurrently mutated gene in melanoma.

To gain a comprehensive view of the genomic landscape in human melanoma tumours, we sequenced the genomes of 25 metastatic melanomas and peripheral blood obtained from the same patients (Supplementary Table 1). Two tumours (ME015 and ME032) were metastases from cutaneous melanomas arising on glabrous (that is, hairless) skin of the extremities, representing the acral subtype. The other tumours were primarily metastases from melanomas originating on hair-bearing skin of the trunk (the most common clinical subtype). Further, ME009 represented a metastasis from a primary melanoma of a patient with a clinical history of chronic ultraviolet exposure.

We obtained 59-fold mean haploid genome coverage for tumour DNA and 32-fold for normal DNA (Supplementary Table 2). On average, 78,775 somatic base substitutions per tumour were identified, consistent with prior reports^{3,4} (Supplementary Table 3). This corresponded to an average mutation rate of 30 per Mb. However, the mutation rate varied by nearly two orders of magnitude across the 25 tumours (Fig. 1). The acral melanomas showed mutation rates

comparable to other solid tumour types (3 and 14 mutations per Mb)^{5,6}, whereas melanomas from the trunk harboured substantially more mutations, in agreement with previous studies^{3,7,8}. In particular, sample ME009 exhibited a striking rate of 111 somatic mutations per Mb, consistent with a history of chronic sun exposure.

In tumours with elevated mutation rates, most nucleotide substitutions were C→T or G→A transitions consistent with ultraviolet irradiation⁹. The variations in mutation rate correlated with differences in the ultraviolet mutational signature. For example, 93% of substitutions in ME009 but only 36% in acral melanoma ME015 were C→T transitions (Fig. 1); these tumours contained the highest and lowest base mutation rates, respectively (111 and 3 mutations per Mb). Interestingly, the acral tumour ME032 also showed a discernible enrichment of ultraviolet-associated mutations (Fig. 1). Thus, genome sequencing readily confirmed the contribution of sun exposure in melanoma aetiology.

In agreement with prior studies^{7,9}, we detected an overall enrichment for dipyrimidines at C→T transitions. Analysis of intragenic C→T mutations yielded a significant bias against such mutations on the transcribed strand for most melanomas, consistent with transcription-coupled repair (Supplementary Fig. 1)^{3,7,10}. Most commonly, C→T mutations occurred at the 3' base of a pyrimidine dinucleotide (CpC or TpC; Supplementary Fig. 2). In contrast, the C→T mutations in sample ME009 (with hypermutation and chronic sun exposure history) more often occurred at the 5' base of a pyrimidine dinucleotide. As expected, the acral tumour ME015 exhibited mutation patterns observed in non-ultraviolet-associated tumour types¹¹, such as an increased mutation rate at CpG dinucleotides relative to their overall genome-wide frequency (Supplementary Fig. 2). These different mutational signatures suggest a complex mechanism of ultraviolet mutagenesis across the clinical spectrum of melanoma, probably reflecting distinct histories of environmental exposures and cutaneous biology.

We detected 9,653 missense, nonsense or splice site mutations in 5,712 genes (out of a total of 14,680 coding mutations; Supplementary Tables 4 and 5), with an estimated specificity of 95% (Supplementary Methods). A mutation of *BRAF*, *BRAF*^{V600E}, was present in 16 of 25 tumours (64%), including the acral melanoma ME015. *NRAS* was mutated in 9 of 25 tumours (36%) in a mutually exclusive fashion with

¹The Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ²Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ³Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Department of Dermatology, University Hospital Essen, D-45122 Essen, Germany. ⁵Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel. ⁶Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁷Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁸Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁹Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. ¹⁰Division of Immunology, Allergy and Infectious Diseases, Department of Dermatology, Medical University of Vienna and CeMM-Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria. ¹¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. †Present address: Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA (M.F.B.); Department of Genomic Medicine, Institute for Applied Cancer Science, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA (T.P.H., Y.L.D., A.P., I.R.W., M.A.S., L.C.).

*These authors contributed equally to this work.

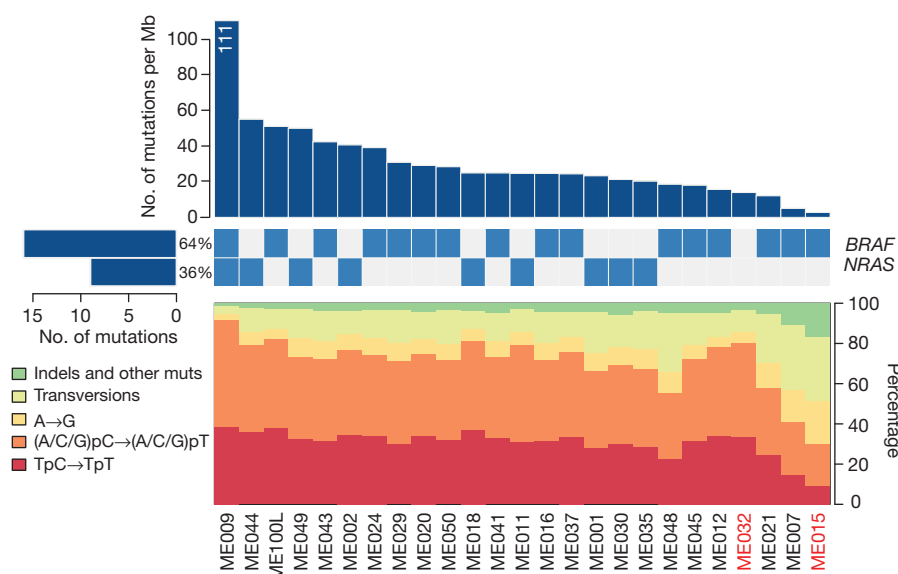


Figure 1 | Elevated mutation rates and spectra indicative of ultraviolet radiation damage. Top bar plot shows somatic mutation rate of 25 sequenced melanoma genomes, in decreasing order. Middle matrix indicates BRAF and NRAS somatic mutation status, with left-adjacent bar plot indicating total

BRAF, with the exception of one non-canonical substitution (*NRAS*^{T501}) in the hypermutated sample ME009. We also identified 6 insertions and 34 deletions in protein coding exons (Supplementary Table 6), including a 21-base-pair (bp) in-frame deletion involving exon 11 of the *KIT* oncogene in the acral tumour ME032 (Supplementary Fig. 3). *KIT* mutations occur in 15% of acral and mucosal melanomas¹², and melanoma patients with activating *KIT* mutations in exon 11 have demonstrated marked responses to imatinib treatment¹³.

We identified an average of 97 structural rearrangements per melanoma genome (range: 6–420) (Supplementary Table 7). In addition to displaying a wide range of rearrangement frequencies, the proportion of intrachromosomal and interchromosomal rearrangements varied widely across genomes. ME029, which harboured the largest number of rearrangements (420), contained only 8 interchromosomal events (Fig. 2a). In contrast, ME020 and ME035 contained 95 and 90 interchromosomal rearrangements, respectively (Fig. 2a). In both cases, the vast majority of interchromosomal rearrangements were restricted to two chromosomes. This pattern is reminiscent of chromothripsis¹⁴, a process involving catastrophic chromosome breakage that has been observed in several tumour types^{15,16}.

106 genes harboured chromosomal rearrangements in two or more samples (Supplementary Table 8). Many recurrently rearranged loci contain large genes or reside at known or suspected fragile sites¹⁷; examples include *FHIT* (six tumours), *MACROD2* (five tumours) and *CSMD1* (four tumours). On the other hand, several known cancer genes were also recurrently rearranged, including the *PTEN* tumour suppressor (four tumours) and *MAGI2* (three tumours), which encodes a protein known to bind and stabilize PTEN. *MAGI2* was also found disrupted in recent whole-genome studies of prostate cancer¹⁸ and a melanoma cell line⁷. Rearrangements involving the 5' untranslated region of the ataxin 2-binding protein 1 gene (*A2BP1*) were observed in 4 tumours. *A2BP1* encodes an RNA binding protein whose genetic disruption has been linked to spinocerebellar ataxia and other neurodegenerative diseases. *A2BP1* undergoes complex splicing regulation in the central nervous system and other tissues¹⁹; in melanoma, these rearrangements may disrupt a known *A2BP1* splice isoform or enable a *de novo* splicing product. Together, these results suggest that chromosomal rearrangements may contribute importantly to melanoma genesis or progression.

An acral melanoma (ME032) harboured the second-largest number of total rearrangements (314; Fig. 2a). We employed high-throughput

number of mutations in each oncogene as well as percent frequency. Bottom plot displays each tumour's somatic mutation spectrum as a percentage of all mutations (right axis). Tumour sample names are indicated at the bottom of the figure, with acral melanomas in red.

PCR followed by massively parallel sequencing to successfully validate 177 of 182 events tested in this sample, confirming its high rate of rearrangement. The elevated frequency of genomic rearrangements in acral melanomas has been reported previously²⁰. In comparison, ME032 exhibited one of the lowest base-pair mutation rates of the melanomas examined (22nd out of 25 samples), suggesting that different tumours might preferentially enact alternative mechanisms of genomic alteration to drive tumorigenesis.

As noted above, many rearrangements in ME032 involved multiple breakpoints within a narrow genomic interval. One such event disrupted the *ETV1* locus. We previously demonstrated an oncogenic role for *ETV1* in melanoma, whose dysregulated expression was associated with upregulation of microphthalmia-associated transcription factor (*MITF*)²¹, the master melanocyte transcriptional regulator and a melanoma lineage survival oncogene²². We validated six distinct rearrangements (four interchromosomal translocations) in ME032 involving breakpoints within *ETV1* introns (Fig. 2b). These events join regions of *ETV1* to distal loci on chromosomes 8, 9, 11 and 15. In support of their possible functional relevance, these rearrangements were associated with high-level *ETV1* amplification in this tumour.

A second complex rearrangement involved the *PREX2* locus. *PREX2* encodes a phosphatidylinositol 3,4,5-trisphosphate RAC exchange factor recently shown to interact with the *PTEN* tumour suppressor and modulate its function². We validated nine somatic rearrangements in the vicinity of *PREX2* (six interchromosomal translocations), including five with intronic breakpoints (Fig. 2c, Supplementary Fig. 4). One event joined specific intronic regions of *PREX2* and *ETV1*. Like *ETV1*, *PREX2* is highly amplified in this tumour, as verified by FISH (fluorescence *in situ* hybridization) analysis (Fig. 2d, Supplementary Fig. 5). The presence of these complex structural rearrangements in addition to amplification may indicate multiple mechanisms of *PREX2* dysregulation in melanoma. More generally, these findings raised the possibility that sites of complex rearrangement might denote genes of functional importance in melanoma.

Next, we calculated the mutational significance of each gene based on the number of mutations detected, gene length and background mutation rates (Table 1, Supplementary Table 9) (see Methods). Eleven genes were found to be significantly mutated across the 25 samples ($Q < 0.01$, where Q is the false discovery rate adjusted P -value). As expected, the two most significant genes were *BRAF* and *NRAS*, mutated in 16 and 9 samples, respectively. Interestingly,

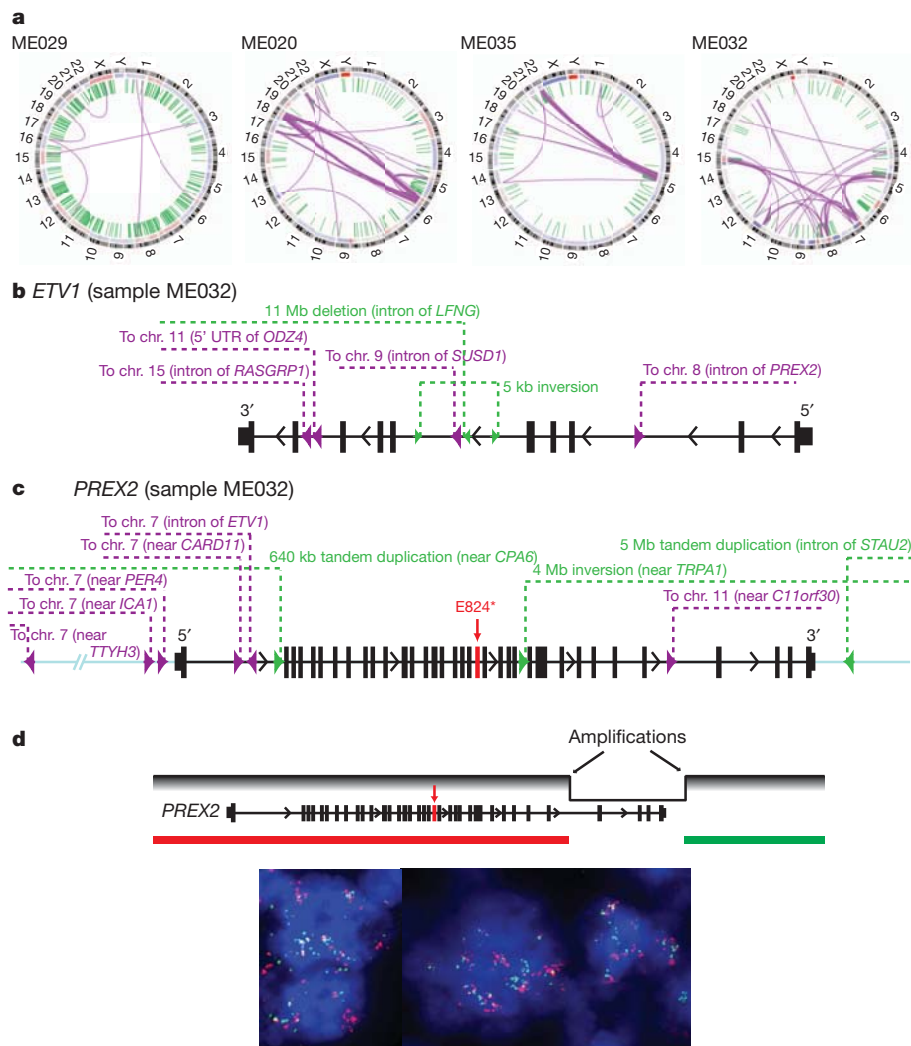


Figure 2 | Hubs of rearrangement breakpoints affect known and putative oncogenes. **a**, Circos plots representing four melanoma genomes with notable structural alterations. Interchromosomal and intrachromosomal rearrangements are shown in purple and green, respectively. **b**, Location of breakpoints associated with *ETV1* in melanoma ME032. **c**, Location of breakpoints associated with *PREX2* in melanoma ME032. The red arrow

indicates a premature stop codon (E824*). All rearrangements in *ETV1* and *PREX2* were validated by high-throughput PCR and deep sequencing. **d**, Confirmation of high-level amplification and rearrangement of *PREX2* in ME032 by dual-colour break-apart FISH. The assignment of red and green FISH probes to the *PREX2* gene region is delineated as bars. Lack of co-localization of red and green probes is indicative of break-apart.

PREX2 scored as one of the top significant genes (Table 1). Furthermore, four samples harboured nonsense truncation mutations in *PREX2*, more than any of the other genes identified as statistically significant in this analysis. *PREX2* mutations have occasionally been reported in colon, lung and pancreatic cancer²³, albeit at low frequencies. Here, we detected 13 non-synonymous point mutations in *PREX2*—including 4 nonsense mutations—and 1 synonymous mutation, with 11 of 25 melanomas harbouring at least 1 non-synonymous mutation. The mutations were distributed throughout the entire length of *PREX2* (Fig. 3a, green circles), and 13 of 14 mutations were non-synonymous, suggestive of positive selection. An analysis of the mutant allele frequencies and estimated tumour purities indicates that at least two mutations are homozygous. One melanoma, ME018, harbours three missense mutations, two of which (I534M and G1581R) appear to co-occur on a single allele based on their observed mutation frequencies. Notably, a *PREX2* nonsense mutation was detected in ME032, in addition to the rearrangements and amplification of this locus present in this tumour (Fig. 2c). This *PREX2* mutation was truncating (E824*), removing the carboxy-terminal region with homology to an inositol phosphatase domain. Based on the allele frequency of this mutation, we infer that it occurs on the

non-amplified allele. Taken together, whole-genome sequencing of this 25-sample discovery cohort identified *PREX2* as a candidate melanoma gene whose amplifications, rearrangements or mutations appeared to undergo positive selection in human melanoma genesis.

To determine the prevalence of *PREX2* mutations in melanoma, we performed bidirectional capillary sequencing in an extension cohort of 107 tumour/normal pairs, comprising 45 tumours and 62 short-term cultures collected from multiple institutions and geographic regions (Supplementary Table 10). We identified 23 somatic base pair mutations and one frame-shift insertion in *PREX2* in this cohort (Fig. 3a; Supplementary Table 11), 15 of which represented non-synonymous changes. We therefore inferred a 14% frequency of non-synonymous *PREX2* mutations in this melanoma cohort.

Discrepant non-synonymous:synonymous ratios were observed between the tumour samples and short-term cultures in the extension cohort. In line with results from the discovery cohort, 100% of *PREX2* mutations detected across 45 tumour samples were non-synonymous in nature ($n = 4$), consistent with positive selection. In contrast, only 55% of the sequence mutations found in the 64 short-term cultures were non-synonymous (a ratio of 11:9). Conceivably, these findings may indicate that subsets of melanoma cells capable of robust growth

Table 1 | Significantly mutated genes in 25 melanoma tumours.

Rank	Gene	Total number of covered bases	Samples with non-synonymous mutations	Non-synonymous mutations	Nonsense mutations	Synonymous mutations	P value	Q value
1	<i>BRAF</i> *	56,520	12	12	0	0	$<10^{-15}$	$<10^{-11}$
2	<i>NRAS</i>	14,160	9	9	0	0	4×10^{-15}	4×10^{-11}
3	<i>MUC4</i>	77,038	19	42	0	17	2×10^{-11}	2×10^{-7}
4	<i>PREX2</i>	127,041	11	13	4	1	2×10^{-8}	8×10^{-5}
5	<i>GOLGA6L6</i>	18,902	5	6	0	1	6×10^{-7}	2×10^{-3}
6	<i>VCX3B</i>	7,132	4	4	0	1	7×10^{-7}	2×10^{-3}
7	<i>POTEH</i>	21,545	5	7	0	1	8×10^{-7}	2×10^{-3}
8	<i>OR2T33</i>	21,978	5	5	0	1	9×10^{-7}	2×10^{-3}
9	<i>C1orf127</i>	53,004	6	6	0	0	2×10^{-6}	3×10^{-3}
10	<i>PRG4</i>	100,212	8	9	0	1	3×10^{-6}	6×10^{-3}
11	<i>MST1</i>	46,400	8	12	0	4	5×10^{-6}	9×10^{-3}

* *BRAF*^{V600E} mutations were detected in four additional samples by exon capture on manual review of Illumina sequencing data (shown in Fig. 1).

in vitro may have experienced reduced selective pressure for *PREX2* mutations. Alternatively (or in addition), the *PREX2* locus may exhibit an enhanced 'local' mutation rate, a by-product of which is the production of variants that undergo positive selection *in vivo*.

To demonstrate the functional relevance of *PREX2* mutations in melanoma tumorigenesis, we ectopically expressed six representative mutations (three truncation variants and three non-synonymous point mutations predicted to carry functional impact²⁴) in TERT-immortalized human melanocytes engineered to express NRAS(G12D) (PMEL-NRAS*)²¹. These melanocytic lines were transplanted into immunodeficient mice alongside control melanocytes expressing either wild-type *PREX2* or GFP (green fluorescent protein). Overexpression of all three truncated variants as well as a point mutant (G844D) of *PREX2* significantly accelerated *in vivo* tumorigenesis when compared to GFP control or wild-type *PREX2*-expressing melanocytes (Fig. 3b, Supplementary Fig. 6). These results therefore affirmed the aforementioned genomic data suggesting that *PREX2* mutations may undergo positive selection *in vivo*. Although the spectrum of *PREX2* mutations in human melanoma (Fig. 3a) is reminiscent of inactivating mutations, our findings suggest that *PREX2* somatic mutations generate truncated or variant proteins that gain oncogenic activity in melanoma cells.

In summary, following recent efforts to characterize whole genomes from several haematologic and solid tumours, we provide the first (to our knowledge) high-resolution view of the genomic landscape across a spectrum of metastatic melanoma tumours. The analysis reveals global

genomic evidence for the role of ultraviolet mutagenesis in melanoma, and identifies several recurrently mutated and rearranged genes not previously implicated in this malignancy. In particular, we discovered that *PREX2* mutations are both recurrent and functionally consequential in melanoma biology. Although its precise mechanism(s) of action remains to be elucidated in melanoma, *PREX2* appears to acquire oncogenic activity through mutations that perturb or inactivate one or more of its cellular functions. This pattern of mutations may exemplify a category of cancer genes that is distinct from 'classic' oncogenes (often characterized by highly recurrent gain-of-function mutations) and tumour suppressors (inactivated by simple loss-of-function alterations). Instead, (over)expression of certain cancer genes with distributed mutation patterns may promote tumorigenicity either through dominant negative effects or more subtle dysregulation of normal protein functions.

Cancer genomics has enabled the discovery and rational application of the first truly effective targeted therapy for metastatic melanoma: *BRAF* mutations predict sensitivity to selective RAF inhibitors^{25–27}. However, the emergence of acquired resistance is rapid and often driven by other genomic events²⁸. Our genomic exploration of the melanoma genomes revealed a large number of complex alterations that probably affect many other genes in addition to *PREX2*. Understanding how this spectrum of genomic aberrations contributes to melanoma genesis and progression should provide new insights into tumour biology, therapeutic resistance and development of treatment regimens aimed at durable control of this malignancy.

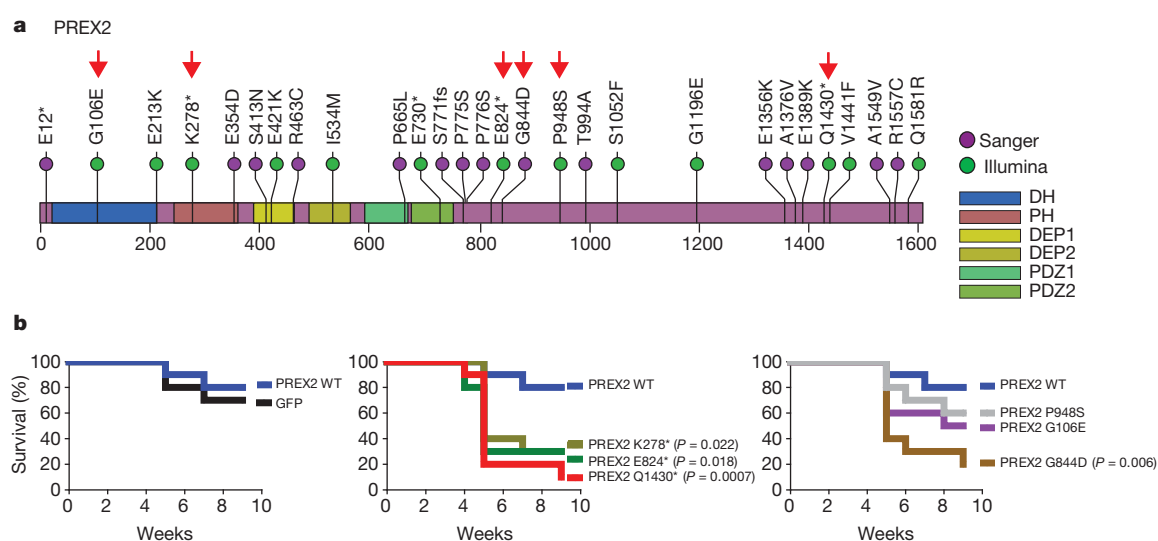


Figure 3 | Mutant *PREX2* expression promotes melanoma genesis. **a**, Non-synonymous sequence mutations detected from Illumina sequencing of 25 melanomas (green) or from capillary sequencing of a validation cohort of 107 additional melanomas (purple). Mutations are dispersed throughout all annotated structural domains of *PREX2*. (DH, DBL homology domain; PH, plekstrin homology domain). The C-terminal half of *PREX2* exhibits

sequence homology to an inositol phosphatase domain. Engineered *PREX2* mutants are labelled with red arrowheads. **b**, Kaplan-Meier curve showing tumour-free survival of NUDE mice (n = 10) injected with PMEL-NRAS* cells expressing GFP, wild-type (left plot), truncated (middle plot) and mutated (right plot) *PREX2* subcutaneously.

METHODS SUMMARY

The complete genomes of 25 metastatic melanomas and patient-matched germline samples were sequenced to approximately 30× and 30× haploid coverage, respectively, on an Illumina GAIIx sequencer (5 cases), and approximately 65× and 32× haploid coverage, respectively, on an Illumina HiSeq 2000 sequencer (20 cases) as paired-end 101-nucleotide reads. Read pairs were aligned to the reference human genome (hg19) using BWA²⁹. Somatic alterations (single base substitutions, small insertions and deletions, and structural rearrangements) were identified according to their presence in the tumour genome and absence from the corresponding normal genome. A subset of rearrangements was validated by PCR and an independent sequencing technology in order to assess the specificity of the detection algorithm. Fluorescence *in situ* hybridization (FISH) was performed to confirm the high level amplification and rearrangement of *PREX2*. Significantly mutated genes were identified by comparing the observed mutations to the background mutation rates calculated for different sequence context categories per tumour sample. 40 exons of *PREX2* were sequenced by PCR and bidirectional capillary sequencing in a validation panel of 107 additional melanoma tumours and short term cultures; mutations were confirmed as somatic by sequencing matched normal DNA. For gain of function studies, *PREX2* mutation constructs were engineered and introduced to PMEL cell lines by lentiviral transduction. To assess the oncogenic roles of *PREX2* mutants, PMEL-NRAS* cells were injected subcutaneously into NUDE mice, and tumour growth was measured over time. A complete description of the materials and methods is provided in Supplementary Information.

Received 7 September 2010; accepted 9 March 2012.

Published online 9 May 2012.

- Chin, L. The genetics of malignant melanoma: lessons from mouse and man. *Nature Rev. Cancer* **3**, 559–570 (2003).
- Fine, B. *et al.* Activation of the PI3K pathway in cancer through inhibition of PTEN by exchange factor P-REX2a. *Science* **325**, 1261–1265 (2009).
- Berger, M. F. *et al.* Integrative analysis of the melanoma transcriptome. *Genome Res.* **20**, 413–427 (2010).
- Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467–472 (2011).
- Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
- Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869–873 (2010).
- Pleasant, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
- Wei, X. *et al.* Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nature Genet.* **43**, 442–446 (2011).
- Drobetsky, E. A., Groszovsky, A. J. & Glickman, B. W. The specificity of UV-induced mutations at an endogenous locus in mammalian cells. *Proc. Natl Acad. Sci. USA* **84**, 9103–9107 (1987).
- Vrieling, H. *et al.* Strand specificity for UV-induced DNA repair and mutations in the Chinese hamster HPRT gene. *Nucleic Acids Res.* **19**, 2411–2415 (1991).
- Rubin, A. F. & Green, P. Mutation patterns in cancer genomes. *Proc. Natl Acad. Sci. USA* **106**, 21766–21770 (2009).
- Curtin, J. A., Busam, K., Pinkel, D. & Bastian, B. C. Somatic activation of KIT in distinct subtypes of melanoma. *J. Clin. Oncol.* **24**, 4340–4346 (2006).
- Hodi, F. S. *et al.* Major response to imatinib mesylate in KIT-mutated melanoma. *J. Clin. Oncol.* **26**, 2046–2051 (2008).
- Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
- Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148**, 59–71 (2012).
- Kloosterman, W. P. *et al.* Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer. *Genome Biol.* **12**, R103 (2011).
- Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898 (2010).
- Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220 (2011).
- Nakahata, S. & Kawamoto, S. Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res.* **33**, 2078–2089 (2005).
- Curtin, J. A. *et al.* Distinct sets of genetic alterations in melanoma. *N. Engl. J. Med.* **353**, 2135–2147 (2005).
- Jané-Valbuena, J. *et al.* An oncogenic role for *ETV1* in melanoma. *Cancer Res.* **70**, 2075–2084 (2010).
- Garraway, L. A. *et al.* Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
- Forbes, S. A. *et al.* COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* **38**, D652–D657 (2010).
- Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
- Flaherty, K. T., Hodi, F. S. & Bastian, B. C. Mutation-driven drug development in melanoma. *Curr. Opin. Oncol.* **22**, 178–183 (2010).
- Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* **363**, 809–819 (2010).
- Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
- Solit, D. B. & Rosen, N. Resistance to BRAF inhibition in melanomas. *N. Engl. J. Med.* **364**, 772–774 (2011).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Illumina sequencing was performed at the Broad Institute and array-based genomic characterization and functional studies were performed at the Belfer Institute of DFCI. We are grateful to the Broad Institute Genome Sequencing Platform, Genome Analysis Platform and Biological Samples Platform. This work was supported by the National Human Genome Research Institute (S.B.G., E.S.L.), National Cancer Institute (M.M., L.C.), FWF-Austrian Science Fund (S.N.W.), NIH Director's New Innovator Award (L.A.G.), Melanoma Research Alliance (L.A.G., L.C.), Starr Cancer Consortium (L.A.G.) and the Burroughs-Wellcome Fund (L.A.G.).

Author Contributions M.F.B., E.H., T.P.H. and Y.L.D. are lead authors; E.S.L., G.G., L.C. and L.A.G. are senior authors. M.F.B. and E.H. performed the genomic analysis of the whole-genome sequencing data. T.P.H. and Y.L.D. performed the molecular biology and mouse experiments to interrogate the function of *PREX2*. E.I., P.G., R.Z. and M.A.S. participated in the functional experiments. M.S.L. performed genomic analysis of mutations and rearrangements. A.P. and X.R. performed FISH studies of *PREX2*. H.Z., K.C., A.Y.S., T.F., S.L.C., Y.D., P.S., D.V., R.J., G.S., A.H.R., T.J.P. and N.S. provided additional computational analyses. A.S. and D.S. contributed melanoma short-term cultures for the extension cohort. I.R.W. contributed clinical information. C.S., R.O., W.W., S.M., D.A. and K.A. participated in DNA sample processing and quality control. E.N. played a project management role. M.F.B., M.S.L., R.O., M.P., L.A., W.W., G.G. and L.A.G. validated candidate rearrangements. J.B. and S.B.G. oversaw the generation of DNA sequence data. J.B., M.M., S.B.G. and T.R.G. contributed to the study design and interpretation of data. S.N.W. contributed most melanoma tumours for the discovery and extension cohort and participated in the study design. J.W. and N.W. also contributed tumour material for the discovery cohort. E.S.L., G.G., L.C. and L.A.G. conceived of and designed the study and participated in the data analysis and interpretation. M.F.B., T.P.H., L.C. and L.A.G. wrote the paper.

Author Information All Illumina sequence data are publicly available in dbGaP (accession number phs000452.v1.p1). Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.C. (LChin@mdanderson.org) or L.A.G. (levi_garraway@dfci.harvard.edu).

Apolipoprotein E controls cerebrovascular integrity via cyclophilin A

Robert D. Bell^{1,2}, Ethan A. Winkler¹, Itender Singh¹, Abhay P. Sagare¹, Rashid Deane¹, Zhenhua Wu¹, David M. Holtzman³, Christer Betsholtz⁴, Annika Armulik^{4,5}, Jan Sallstrom¹, Bradford C. Berk² & Berislav V. Zlokovic^{1,6}

Human apolipoprotein E has three isoforms: APOE2, APOE3 and APOE4¹. APOE4 is a major genetic risk factor for Alzheimer's disease^{2,3} and is associated with Down's syndrome dementia and poor neurological outcome after traumatic brain injury and haemorrhage³. Neurovascular dysfunction is present in normal APOE4 carriers^{4–6} and individuals with APOE4-associated disorders^{3,7–10}. In mice, lack of *Apoe* leads to blood–brain barrier (BBB) breakdown^{11,12}, whereas APOE4 increases BBB susceptibility to injury¹³. How APOE genotype affects brain microcirculation remains elusive. Using different APOE transgenic mice, including mice with ablation and/or inhibition of cyclophilin A (CypA), here we show that expression of APOE4 and lack of murine *Apoe*, but not APOE2 and APOE3, leads to BBB breakdown by activating a proinflammatory CypA–nuclear factor- κ B–matrix-metalloproteinase-9 pathway in pericytes. This, in turn, leads to neuronal uptake of multiple blood-derived neurotoxic proteins, and microvascular and cerebral blood flow reductions. We show that the vascular defects in *Apoe*-deficient and APOE4-expressing mice precede neuronal dysfunction and can initiate neurodegenerative changes. Astrocyte-secreted APOE3, but not APOE4, suppressed the CypA–nuclear factor- κ B–matrix-metalloproteinase-9 pathway in pericytes through a lipoprotein receptor. Our data suggest that CypA is a key target for treating APOE4-mediated neurovascular injury and the resulting neuronal dysfunction and degeneration.

Astrocytes are a major source of APOE in the brain². Importantly, astrocyte-secreted molecules transduce signals to brain microvessels acting on pericytes^{7,14,15}. To understand the effects of APOE on brain microcirculation we studied mice with targeted replacement of murine *Apoe* with each human APOE isoform (TR-APOE)¹⁶, mice lacking murine *Apoe* (*Apoe*^{−/−}), mice expressing each human APOE isoform under control of the astrocyte-specific glial fibrillary acidic protein (GFAP) promoter on an *Apoe*-null background, and *Apoe*^{−/−} and APOE4 transgenic mice with ablation and/or pharmacological inhibition of CypA (see Supplementary Information). In search of molecules that could mediate BBB dysfunction in *Apoe*^{−/−} and APOE4 mice, we focused on the proinflammatory cytokine CypA, previously demonstrated to have deleterious effects on the vascular system in *Apoe*^{−/−} mice with aortic aneurysms and atherosclerosis^{17,18}.

Using multiphoton microscopy of tetramethylrhodamine-conjugated dextran (TMR-dextran)¹⁴, we show an intact BBB in TR-APOE2 and TR-APOE3 mice and a leaky BBB in TR-APOE4 and *Apoe*^{−/−} mice (Fig. 1a and Supplementary Fig. 1a, b), suggesting that APOE2, APOE3 and murine *Apoe* effectively maintain the BBB, whereas APOE4 promotes BBB disruption. These findings have been replicated in mice expressing each human APOE isoform under control of the GFAP promoter (not shown). Notably, genetic ablation of CypA (encoded by *Ppia*) eliminated BBB damage in *Apoe*^{−/−} *Ppia*^{−/−} mice (Fig. 1a and Supplementary Fig. 1a, b).

Compared to littermate controls, TR- or GFAP-APOE2 and -APOE3 mice, *Apoe*^{−/−} and APOE4 mice had five- to sixfold higher CypA levels in cerebral microvessels (Fig. 1b and Supplementary Fig. 1c, d), mainly because of an increased CypA expression in pericytes (Fig. 1c and Supplementary Fig. 1e). CypA levels in microvessel-depleted brain were not affected by APOE (Supplementary Fig. 1f). These data suggest that APOE2, APOE3 and murine *Apoe*, but not APOE4, effectively maintain physiological CypA levels in brain microvessels by controlling CypA expression in pericytes. To determine whether BBB disruption in APOE4 mice can be corrected with cyclosporine A, a drug that binds intracellular CypA and inhibits its effects¹⁹, we treated TR-APOE4 and GFAP-APOE4 mice with a low dose of cyclosporine A previously shown not to cause systemic or central nervous system toxicity. Cyclosporine A accumulates in brain microvessels, but does not cross the BBB²⁰. In APOE4 mice cyclosporine A eliminated BBB disruption (Fig. 1a and Supplementary Fig. 1a, b) and neuronal accumulation of systemically administered cadaverine¹⁵ (Supplementary Fig. 1g), indicating that BBB changes are reversible and CypA can be therapeutically targeted to correct APOE4-induced BBB breakdown.

To understand better the pathological implications of BBB breakdown, we studied leakage of endogenous blood-derived proteins in the brain. As shown in the hippocampus, 18-month-old GFAP-APOE3 and control mice had negligible extravascular accumulation of serum IgG in contrast to GFAP-APOE4 and *Apoe*^{−/−} mice (Supplementary Fig. 2a, b). *Ppia* genetic deletion eliminated IgG extravascular deposits (Supplementary Fig. 2a, b) and neuronal accumulation in *Apoe*^{−/−} and APOE4 mice (Fig. 1d). Cyclosporine A diminished IgG leakage by ~80% in TR-APOE4 or GFAP-APOE4 mice (Supplementary Fig. 2c) and inhibited neuronal accumulation of blood-derived thrombin and fibrin (Fig. 1e), consistent with restoration of the BBB. APOE4 mice had numerous brain perivascular fibrin and haemosiderin foci (Supplementary Fig. 2d–f) and elevated thrombin levels that were normalized with cyclosporine A (Supplementary Fig. 2g, h). Thrombin is neurotoxic²¹, fibrin accelerates neurovascular damage²² and haemosiderin generates reactive oxygen species²³, thus implicating multiple potential BBB-derived sources of injury.

To elucidate the molecular mechanisms underlying CypA-mediated BBB breakdown we studied matrix metalloproteinases (MMP)2 and MMP9 (gelatinases), which are activated by CypA in the vessel wall in a mouse model of aortic aneurysm¹⁷. Gelatinases disrupt the BBB by degrading the capillary basement membrane and tight-junction proteins^{7,24}. Multiphoton microscopy of DQ-gelatin²⁵ revealed an increase in cerebrovascular gelatinase activity in *Apoe*^{−/−} *Ppia*^{+/+} and TR-APOE4 mice compared with controls, TR-APOE2 and TR-APOE3 mice (Fig. 2a, b). Gelatin zymography of brain tissue demonstrated an increase in pro-MMP9 and activated MMP9, but not MMP2, in *Apoe*^{−/−} and TR-APOE4 mice (Fig. 2c), which correlated with the

¹Center for Neurodegenerative and Vascular Brain Disorders, University of Rochester Medical Center, Rochester, New York 14642, USA. ²Aab Cardiovascular Research Institute, University of Rochester Medical Center, Rochester, New York 14642, USA. ³Department of Neurology, Hope Center for Neurological Disorders, Knight Alzheimer's Disease Research Center, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. ⁴Division of Vascular Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm SE-171 77, Sweden. ⁵Institute of Neuropathology, University Hospital Zürich, CH-8091, Zürich, Switzerland. ⁶Center for Neurodegeneration and Regeneration, Zilkha Neurogenetic Institute and Department of Physiology and Biophysics, University of Southern California, Keck School of Medicine, Los Angeles, California 90089, USA.

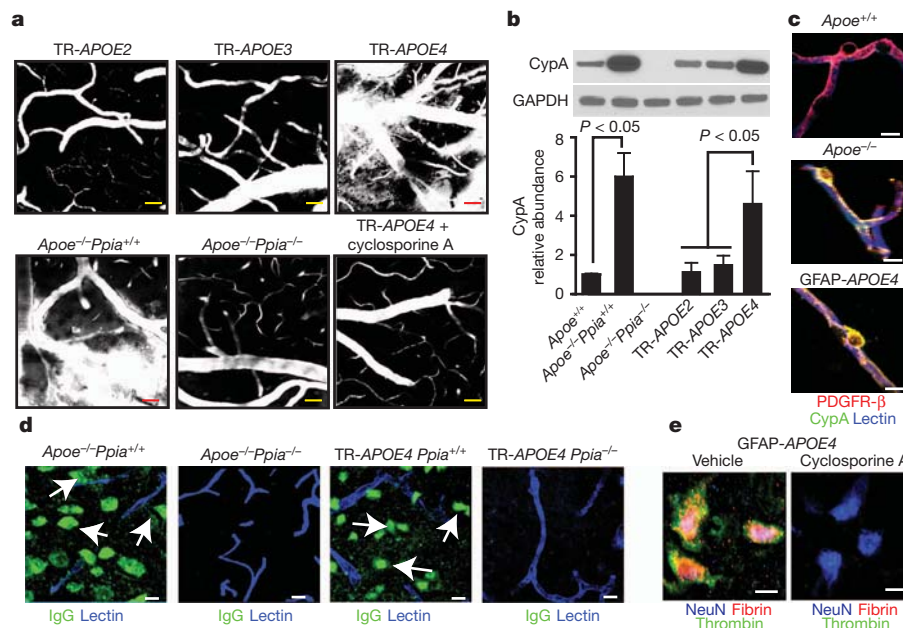
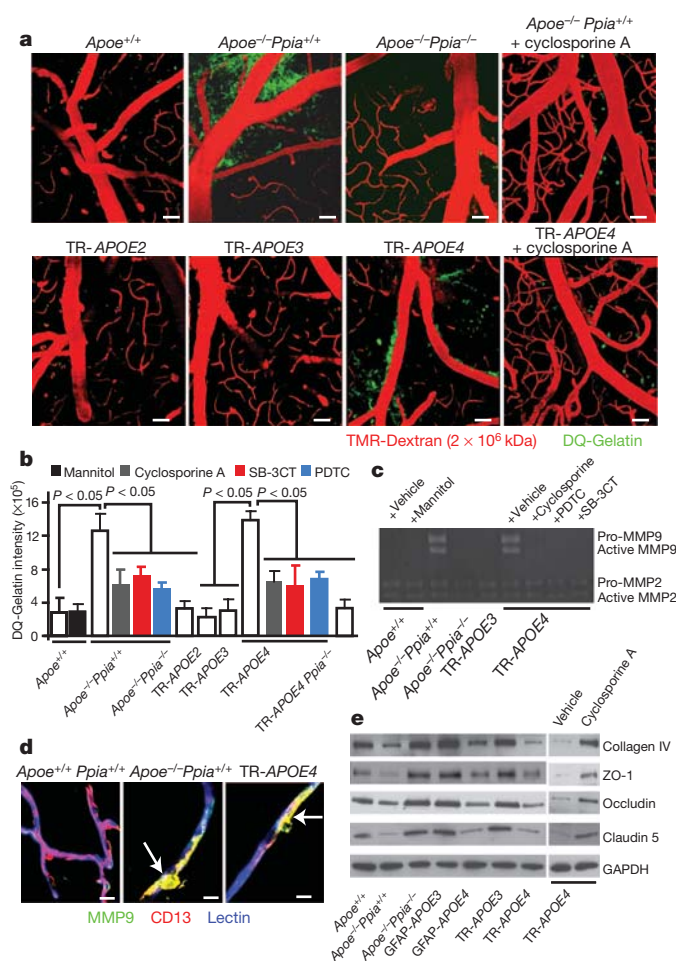


Figure 1 | CypA deficiency or inhibition reverses BBB breakdown in *ApoE*^{-/-} and *APOE4* mice. **a**, Multiphoton microscopy of TMR-Dextran (white) in 6-month-old TR-APOE2, TR-APOE3, TR-APOE4, *ApoE*^{-/-} *Ppia*^{+/+}, *ApoE*^{-/-} *Ppia*^{-/-} and cyclosporine A-treated TR-APOE4 mice. Scale bar, 20 μ m. **b**, CypA immunoblotting in brain microvessels from APOE transgenic mice. GAPDH, Glyceraldehyde 3-phosphate dehydrogenase. **c**, CypA (green) co-localization with platelet-derived growth factor receptor beta (PDGFR- β)-positive pericytes (red; yellow indicates merged) in

hippocampal microvessels from *ApoE*^{+/+}, *ApoE*^{-/-} and GFAP-APOE4 mice. Blue indicates lectin-positive endothelium. Scale bar, 10 μ m. **d**, IgG neuronal uptake (green; blue indicates lectin-positive vessels) in *ApoE*^{-/-} *Ppia*^{+/+}, *ApoE*^{-/-} *Ppia*^{-/-}, TR-APOE4 *Ppia*^{+/+} and TR-APOE4 *Ppia*^{-/-} mice. Scale bar, 10 μ m. **e**, Fibrin (red) and thrombin (green) in neuronal nuclear antigen (NeuN)-positive neurons (blue) in the hippocampus of 9-month-old cyclosporine A-treated and -untreated GFAP-APOE4 mice. **a** and **c–e** are representative of results from 4–6 experiments. Scale bar, 10 μ m.



appearance of MMP9-positive pericytes (Fig. 2d and Supplementary Fig. 3a, b). To establish causality and demonstrate that increased MMP9 activity does not only correlate with, but is required for, BBB breakdown in *ApoE*^{-/-} and TR-APOE4 mice, we studied the effects of pharmacological inhibition of MMP9 *in vivo* with 2-[[[(4-phenoxyphenyl) sulfonyl]methyl]-Thiirane (SB-3CT), an MMP9 inhibitor, and of MMP9 and MMP2 silencing by short interfering (si)RNA administration into the hippocampus, as reported²⁶. SB-3CT eliminated MMP9 gelatinase activity (Fig. 2b, c) and reversed the leaky BBB phenotype (Supplementary Fig. 4a) in both mouse lines. Similarly, MMP9, but not MMP2, silencing reversed the BBB phenotype in TR-APOE4 mice (Supplementary Fig. 4b). Similar results were obtained with *siMmp9* treatment in *ApoE*^{-/-} mice (not shown).

Consistent with MMP9 activation, several MMP9 substrates including collagen IV and tight-junction proteins ZO-1 (also known as Tjp1), occludin and claudin 5, which are required for normal BBB integrity^{7,24}, were reduced in brain microvessels in young *ApoE*^{-/-} and APOE4

Figure 2 | CypA activates the NF- κ B-MMP9 pathway causing BBB breakdown in *ApoE*^{-/-} and *APOE4* mice. **a**, Multiphoton microscopy of DQ-gelatin (green) in 8–9-month-old control (*ApoE*^{+/+}), *ApoE*^{-/-} *Ppia*^{+/+}, *ApoE*^{-/-} *Ppia*^{-/-}, TR-APOE2, TR-APOE3, TR-APOE4 and cyclosporine A-treated *ApoE*^{-/-} *Ppia*^{+/+} and TR-APOE4 mice. Red indicates cortical vessels. Scale bar, 25 μ m. **b**, Quantification of DQ-gelatin signal in APOE transgenic mice and effects of cyclosporine A, PDTC, SB-3CT and CypA deletion in *ApoE*^{-/-} and TR-APOE4 mice. Mean \pm s.e.m., $n = 3–6$ animals per group. **c**, Gelatin zymography of brain tissue in control (*ApoE*^{+/+}), *ApoE*^{-/-} *Ppia*^{+/+}, *ApoE*^{-/-} *Ppia*^{-/-}, TR-APOE3 and TR-APOE4 mice treated with vehicle, cyclosporine A, PDTC or SB-3CT. **d**, MMP9 (green) co-localization with CD13-positive pericytes (red; yellow indicates merged) in cortical microvessels from 9-month-old *ApoE*^{+/+} *Ppia*^{+/+}, *ApoE*^{-/-} *Ppia*^{+/+} and TR-APOE4 mice. Blue indicates lectin-positive endothelium. Scale bar, 10 μ m. **e**, Reduced collagen IV, ZO-1, occludin and claudin 5 levels in 2-week-old *ApoE*^{-/-} and TR-APOE4 mice and reversal by CypA ablation (*ApoE*^{-/-} *Ppia*^{+/+}) and cyclosporine A (TR-APOE4). **c** and **e** are representative of results from 4–6 experiments.

mice, indicating BBB breakdown (Fig. 2e and Supplementary Fig. 3c–f). SB-3CT (Supplementary Fig. 4c) and *siMMP-9*, but not *siMMP-2* or control siRNA, normalized the levels of tight-junction and basement-membrane proteins in *APOE4* (Supplementary Fig. 4b) and *Apoe*^{−/−} mice (not shown). These data explain how MMP9 inhibition permits reversal of BBB disruption in *Apoe*^{−/−} and *APOE4* mice at a molecular level. Notably, genetic deletion of *Ppia* or cyclosporine A substantially inhibited gelatinase/MMP9 activity (Fig. 2a–c) and restored the basement-membrane and tight-junction proteins in *Apoe*^{−/−} *Ppia*^{−/−} and TR-*APOE4* mice (Fig. 2e and Supplementary Figs 3c–f, 4c).

Nuclear-factor-κB (NF-κB) transcriptionally activates MMP9 in cerebral vessels, causing BBB breakdown²⁴. Consistent with findings that CypA at pathophysiological levels activates NF-κB and the NF-κB–MMP9 pathway^{18,24,27}, we found NF-κB nuclear translocation in brain capillary pericytes in *Apoe*^{−/−} and *APOE4* mice (Supplementary Fig. 5a, b). In both *Apoe*^{−/−} and *APOE4* mice, NF-κB nuclear translocation was inhibited by *Ppia* gene deletion and/or cyclosporine A (Supplementary Fig. 5a, b).

To establish the causality between NF-κB activation, increased gelatinase activity and BBB breakdown, we treated *Apoe*^{−/−} and *APOE4* mice with pyrrolidine dithiocarbamate (PDTC), an inhibitor of NF-κB nuclear translocation. In *Apoe*^{−/−} and *APOE4* mice, PDTC markedly reduced MMP9 activation (Fig. 2b, c and Supplementary Fig. 3b) and reversed the leaky BBB phenotype (Supplementary Fig. 4a). Consistent with these data, NF-κB inhibition in the hippocampus by silencing *Rela*, which encodes the p65 subunit of NF-κB, inhibited MMP9 and reversed the BBB phenotype in *APOE4* mice (Supplementary Fig. 4c). Similar results were obtained by silencing *Rela* in *Apoe*^{−/−} mice (not shown). Our findings therefore clearly establish that each of the molecules studied (that is, CypA, NF-κB and MMP9) have important and required roles in BBB disruption in *Apoe*^{−/−} and *APOE4* mice.

Consistent with reports that chronic BBB breakdown leads to microvascular reductions^{7,14,22}, we found microvascular degeneration in *APOE4* and *Apoe*^{−/−} mice, including DNA fragmentation in pericytes and endothelial cells (Supplementary Fig. 6a), diminished pericyte coverage (Supplementary Fig. 5b, c) and reductions in microvascular length (Fig. 3a and Supplementary Fig. 6d), which correlated with the degree of BBB breakdown (Supplementary Fig. 6e) and regional cerebral blood flow (CBF) reductions (Fig. 3b and Supplementary Table 1). Notably, *Ppia* deletion and cyclosporine A, SB-3CT or PDTC normalized microvascular reductions in *Apoe*^{−/−} and *APOE4* mice (Fig. 3a and Supplementary Fig. 6c). *Ppia* deletion also normalized CBF reductions in *Apoe*^{−/−} mice (Fig. 3b).

Given that CypA expression and NF-κB and MMP9 activation are increased in pericytes in *Apoe*^{−/−} and *APOE4* mice, we next studied which low-density lipoprotein (LDL)/APOE receptor¹ in pericytes regulates CypA in response to astrocyte-derived APOE. After confirming the specificity of our siRNA reagents (Supplementary Fig. 7), we showed by quantifying the effects of siRNA inhibition (Fig. 4a and Supplementary Fig. 8a–d) and by administering antibodies to block the function of specific LDL receptors (Supplementary Fig. 8e–g) that astrocyte-derived APOE3 and murine *Apoe* require low-density lipoprotein receptor-related protein 1 (LRP1) to maintain CypA synthesis within a physiological range. Adenoviral-mediated re-expression of a human *LRP1* minigene rescued the ability of APOE3 to downregulate *Ppia* mRNA (Fig. 4a) and protein (Supplementary Fig. 8b) in pericytes with siRNA-induced LRP1 knockdown. By imaging APOE/LRP1 proximity ligation in pericytes we demonstrated that APOE3 (Fig. 4b) and murine *Apoe* (not shown) bind with high affinity to LRP1, whereas the APOE4–LRP1 interaction was barely detectable (Fig. 4b). Together these data explain at the molecular level why APOE4 is unable to properly regulate physiologic CypA levels, which is consistent with previously reported statistically insignificant interactions of APOE4 with LRP1 in cerebral microvessels and at the BBB *in vivo*²⁸.

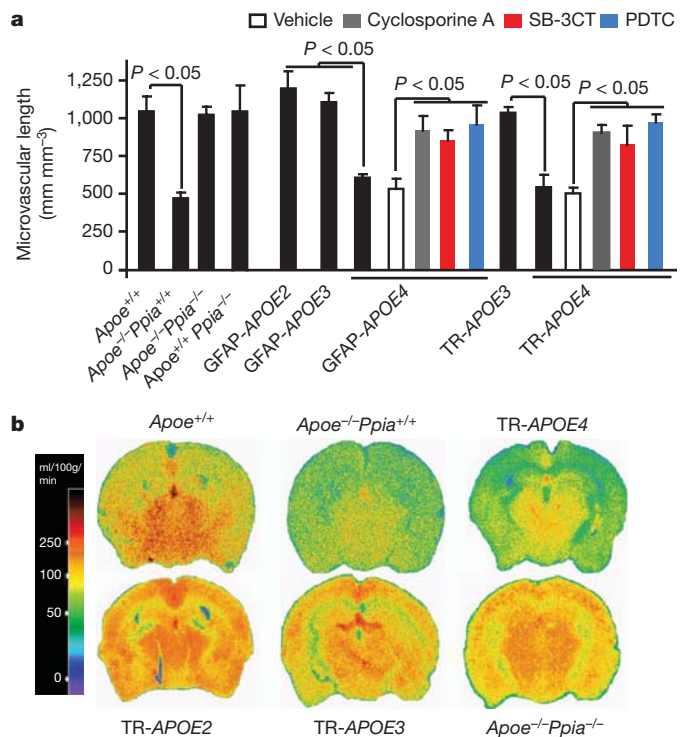


Figure 3 | CypA ablation or inhibition reverses microvascular and CBF reductions in *Apoe*^{−/−} and *APOE4* mice. **a**, Capillary length in the hippocampus of APOE transgenic mice including *Apoe*^{−/−} *Ppia*^{+/+}, *Apoe*^{−/−} *Ppia*^{−/−} and GFAP-*APOE4* and TR-*APOE4* mice treated with cyclosporine A, SB-3CT or PDTC (mean ± s.e.m., *n* = 5 animals per group). **b**, ¹⁴C-iodoantipyrine CBF autoradiograms in 9-month-old transgenic APOE mice. **b** is representative of results from 6 experiments.

Ppia silencing, cyclosporine A and astrocyte-derived APOE3, but not APOE4, inhibited NF-κB nuclear translocation in *Apoe*^{−/−} pericytes (Supplementary Fig. 9a), as *in vivo*. By using *siLrp1* silencing, cyclosporine A or PDTC, we showed that LRP1 is required for APOE3-mediated inhibition of NF-κB-dependent MMP9 activation and transcriptional suppression²⁴ (Supplementary Fig. 9). APOE4 did not have an effect on MMP9 in pericytes, consistent with its barely detectable binding to LRP1 (Fig. 4b). *In vivo*, LRP1 inhibition in APOE3 mice through siRNA administration in the hippocampus²⁶ reproduced vascular phenotypes seen in APOE4 mice, including elevated CypA and MMP9 levels and increased CypA and MMP9 expression in pericytes (Fig. 4c and Supplementary Fig. 10a, b), and BBB breakdown (Fig. 4d). As expected, LRP1 inhibition in mice with *Ppia* genetic deletion did not influence MMP9 expression in pericytes (Supplementary Fig. 10d) or BBB integrity (Supplementary Fig. 10d). Together, these data clearly implicate APOE3/LRP1-mediated CypA regulation in pericytes, conferring APOE3 isoform-specific protection of the BBB (Fig. 4e).

Vascular defects in *Apoe*^{−/−} and APOE4 mice were detectable at 2 weeks of age, including leakage of dextran (Fig. 5a) and serum IgG (Supplementary Fig. 11a), and reductions in tight-junction and basement-membrane proteins, pericyte coverage, capillary length and regional CBF, which progressively increased with age (Fig. 2e and Supplementary Figs 11b–e, 12 and Supplementary Table 1). We next asked whether vascular damage precedes neuronal changes in *Apoe*^{−/−} mice²⁹, and neuronal and synaptic dysfunction in APOE4 mice². Cortical activity determined *in vivo* by voltage-sensitive dye (VSD) imaging indicated normal time-lapse imaging profiles in 2-week-old *Apoe*^{−/−} and APOE4 mice (Fig. 5b and Supplementary Fig. 13a, b) and normal neuritic density and levels of pre-synaptic and post-synaptic proteins (Supplementary Fig. 14a–d). At 4 months of age, however, *Apoe*^{−/−} and APOE4 mice showed a lower amplitude

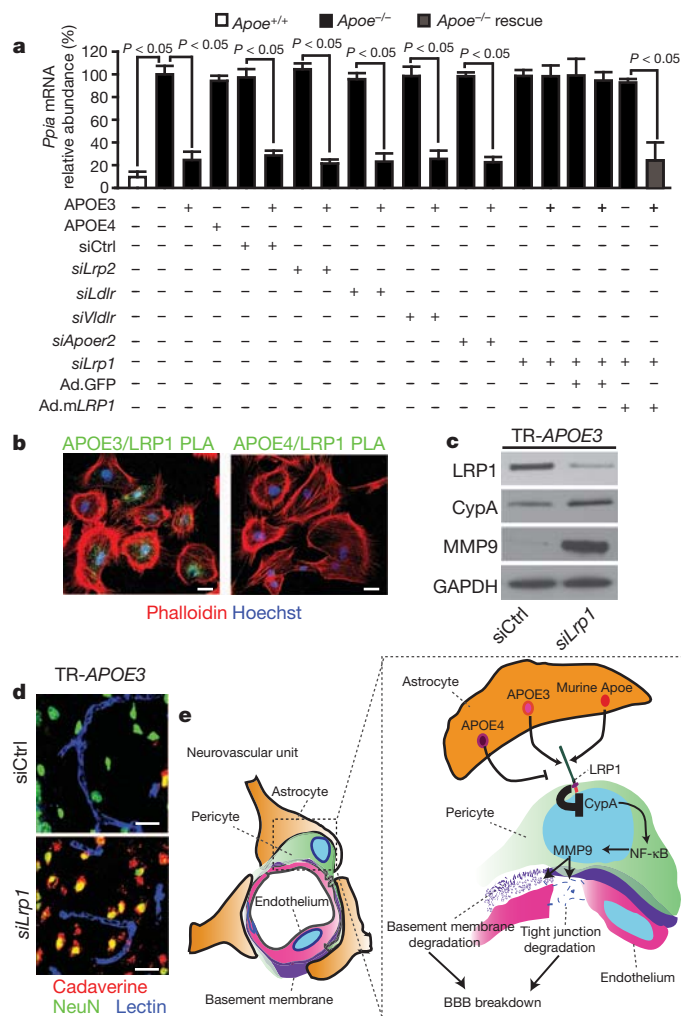


Figure 4 | APOE isoform-specific regulation of CypA-NF- κ B-MMP9 pathway in pericytes. **a**, $Ppia$ mRNA quantification in $ApoE^{-/-}$ pericytes after treatment with astrocyte-secreted APOE3, APOE4, siRNA silencing of low-density lipoprotein receptor-related protein 2 (*Lrp2*), low-density lipoprotein receptor (*Ldlr*), very-low-density lipoprotein receptor (*Vldlr*), apolipoprotein E receptor 2 (*Apoer2*), *Lrp1* and adenoviral-mediated re-expression of *LRP1* minigene (Ad.mLRP1). GFP, green fluorescent protein. Mean \pm s.e.m., $n = 3$ independent cultures. **b**, Proximity ligation assay (PLA) of APOE3 and APOE4 interaction with LRP1 (green). Red and blue signals indicate phalloidin and nuclear staining, respectively. Scale bar, 10 μ m. **c,d**, LRP1, CypA and MMP9 immunodetection (**c**) and neuronal uptake (NeuN, green) of cadaverine-Alexa-Fluor-555 (red; yellow indicates merged) (**d**) in the hippocampus of 6-month-old GFAP-APOE3 mice after *siLrp1* or control siRNA (*siCtrl*) infusion. Blue indicates lectin-positive capillaries. Scale bar, 25 μ m. **e**, A schematic showing that astrocyte-secreted APOE3 and murine ApoE, but not APOE4, signal to pericytes via LRP1, suppressing the CypA-NF- κ B-MMP9 pathway that causes BBB breakdown. **b-d**, Data are representative of results from 6 experiments.

of the VSD signal, longer time-to-peak and a slower duration of the spreading of depolarization (Fig. 5c, d), which was accompanied by age-dependent reductions in neuritic density and pre-synaptic and post-synaptic proteins (Supplementary Fig. 14a-f). These data indicate that $ApoE^{-/-}$ and APOE4 mice develop vascular defects before neuronal and synaptic changes occur. Cyclosporine A, PDTC and SB-3CT improved functional and structural neuronal changes in $ApoE^{-/-}$ and APOE4 mice (Fig. 5e and Supplementary Fig. 15), indicating that normalization of BBB integrity through inhibition of the CypA-NF- κ B-MMP9 pathway is required for neuronal and synaptic repair.

Understanding the contribution of APOE4 to the pathogenesis of Alzheimer's disease may be one of the most important avenues to a

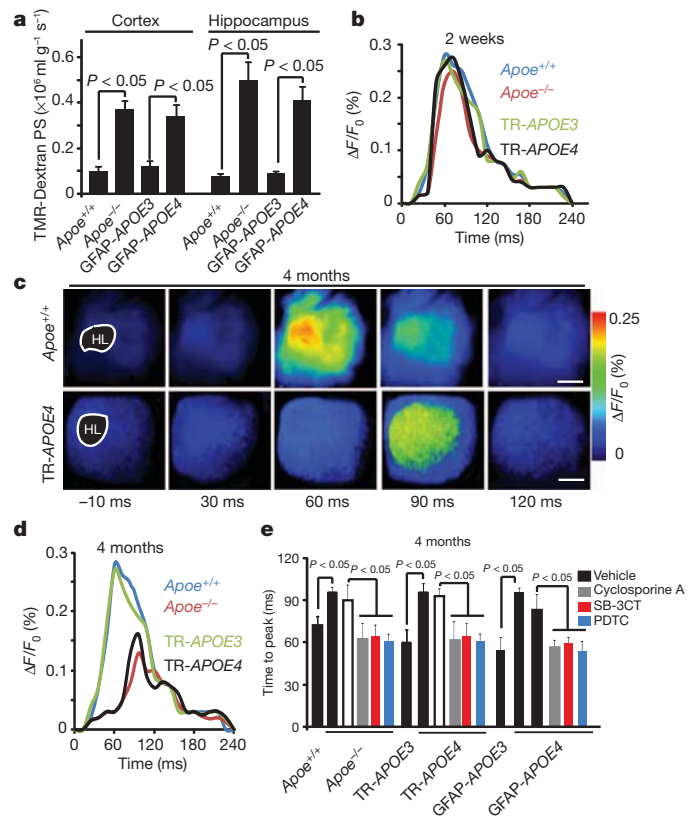


Figure 5 | Vascular defects in $ApoE^{-/-}$ and APOE4 mice precede neuronal dysfunction. **a**, The BBB permeability surface (PS) product for TMR-dextran (40,000 Da) in the cortex and hippocampus of 2-week-old $ApoE^{+/+}$, $ApoE^{-/-}$, GFAP-APOE3 and GFAP-APOE4 mice measured by noninvasive fluorescence spectroscopy. Data are mean \pm s.e.m., $n = 5$ animals per group. **b**, Representative time-lapse-imaging profile analysis of VSD signal response in the hind-limb (HL) somatosensory cortex after stimulation in 2-week-old $ApoE^{+/+}$, $ApoE^{-/-}$, TR-APOE3 and TR-APOE4 mice. $\Delta F/F_0$ indicates the percentage change in fluorescence from the baseline fluorescent signal following stimulation. **c**, VSD imaging of cortical responses to hind-limb stimulation in 4-month-old $ApoE^{+/+}$ and TR-APOE4 mice. Scale bar, 500 μ m. **d**, Representative VSD signal responses in the hind-limb somatosensory cortex region after stimulation in 4-month-old $ApoE^{+/+}$, $ApoE^{-/-}$, TR-APOE3 and TR-APOE4 mice. **e**, Time-to-peak in fluorescent VSD signal after hind-limb stimulation in 4-month-old $ApoE^{+/+}$, $ApoE^{-/-}$, TR-APOE3, TR-APOE4, GFAP-APOE3 and GFAP-APOE4 mice and in $ApoE^{-/-}$, TR-APOE4 and GFAP-APOE4 mice treated with cyclosporine A, SB-3CT, PDTC or vehicle. Data are mean \pm s.e.m., $n = 5$ animals per group.

new therapy. Neurovascular dysfunction and BBB defects have been shown in Alzheimer's disease^{7,30}. The findings from this study indicating that abnormal vessels and pericytes can be involved provide an alternative way of thinking about Alzheimer's disease and neurological disorders affected by APOE4. Our findings demonstrate that APOE maintains cerebrovascular integrity necessary for normal neuronal function by regulating the CypA-NF- κ B-MMP9 pathway in pericytes in an isoform-specific manner (Fig. 4e). We also show that CypA is a key target for treating APOE4-mediated neurovascular defects and the resulting neuronal dysfunction.

METHODS SUMMARY

Animals. $ApoE^{-/-}$, GFAP-APOE mice on murine $ApoE$ -null background and $Ppia^{-/-}$ mice were acquired from Jackson Laboratories. TR-APOE mice were generated as previously described¹⁶. The $Ppia^{-/-}$ mice were crossed with the $ApoE^{-/-}$ and TR-APOE4 mice to generate the $ApoE^{-/-}$ $Ppia^{-/-}$ and TR-APOE4 $Ppia^{-/-}$ mice used in the present study, respectively. Mice were housed in plastic cages on a 12-hour light cycle with ad libitum access to water and a standard laboratory diet. All studies were performed in accordance with the University of Rochester Institutional Animal Care and Use Committee using

National Institutes of Health guidelines. All lines were maintained on a C57BL/6 background. No phenotypic differences were found between littermate control animals. Please see Supplementary Information for discussion on lipid profile and physiological parameters.

Pharmacological inhibition. In some studies, *APOE4* or *ApoE*^{-/-} mice were treated for 7 consecutive days with a low intraperitoneal nontoxic dose of cyclosporine A (Sigma, 30024-25; 10 mg kg⁻¹ day⁻¹ for 3 days followed by 5 mg kg⁻¹ day⁻¹ for 4 days), or PDTC (100 mg kg⁻¹ day⁻¹) or SB-3CT (25 mg kg⁻¹ day⁻¹).

In vivo siRNA infusion. siRNA-mediated *Lrp1*, *Mmp9*, *Mmp2* and *Rela* knockdown was performed as previously described²⁶.

BBB permeability assays. *In vivo* multiphoton imaging of TMR-dextran and detection of endogenous IgG, fibrin, thrombin and Prussian blue deposits in brain tissue was performed as previously described¹⁴. Detection of neuronal uptake of systemically administered Alexa fluor 555-conjugated cadaverine was performed as described¹⁵.

Statistical analysis. Data were analysed by multifactorial analysis of variance (ANOVA) followed by Tukey post-hoc tests and Pearson's correlation analysis using GraphPad Prism 3.0 software. A *P* value of less than 0.05 was considered statistically significant in all studies.

A complete description of all experiments performed and associated references are available in the Supplementary Information section.

Received 24 October 2011; accepted 26 March 2012.

Published online 16 May 2012.

- Mahley, R. W., Weisgraber, K. H. & Huang, Y. Apolipoprotein E: structure determines function, from atherosclerosis to Alzheimer's disease to AIDS. *J. Lipid Res.* **50** (Suppl.), S183–S188 (2009).
- Kim, J., Basak, J. M. & Holtzman, D. M. The role of apolipoprotein E in Alzheimer's disease. *Neuron* **63**, 287–303 (2009).
- Verghese, P. B., Castellano, J. M. & Holtzman, D. M. Apolipoprotein E in Alzheimer's disease and other neurological disorders. *Lancet Neurol.* **10**, 241–252 (2011).
- Thambisetty, M., Beason-Held, L., An, Y., Kraut, M. A. & Resnick, S. M. APOE ε4 genotype and longitudinal changes in cerebral blood flow in normal aging. *Arch. Neurol.* **67**, 93–98 (2010).
- Sheline, Y. I. et al. APOE4 allele disrupts resting state fMRI connectivity in the absence of amyloid plaques or decreased CSF Aβ42. *J. Neurosci.* **30**, 17035–17040 (2010).
- Reiman, E. M. et al. Functional brain abnormalities in young adults at genetic risk for late-onset Alzheimer's dementia. *Proc. Natl Acad. Sci. USA* **101**, 284–289 (2004).
- Zlokovic, B. V. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nature Rev. Neurosci.* **12**, 723–738 (2011).
- Snowdon, D. A. et al. Brain infarction and the clinical expression of Alzheimer disease. The Nun Study. *J. Am. Med. Assoc.* **277**, 813–817 (1997).
- Vermeer, S. E. et al. Silent brain infarcts and the risk of dementia and cognitive decline. *N. Engl. J. Med.* **348**, 1215–1222 (2003).
- Ruitenberg, A. et al. Cerebral hypoperfusion and clinical onset of dementia: the Rotterdam Study. *Ann. Neurol.* **57**, 789–794 (2005).
- Methia, N. et al. ApoE deficiency compromises the blood–brain barrier especially after injury. *Mol. Med.* **7**, 810–815 (2001).
- Hafezi-Moghadam, A., Thomas, K. L. & Wagner, D. D. ApoE deficiency leads to a progressive age-dependent blood–brain barrier leakage. *Am. J. Physiol. Cell Physiol.* **292**, C1256–C1262 (2007).
- Nishitsuji, K., Hosono, T., Nakamura, T., Bu, G. & Michikawa, M. Apolipoprotein E regulates the integrity of tight junctions in an isoform-dependent manner in an *in vitro* blood–brain barrier model. *J. Biochem.* **286**, 17536–17542 (2011).
- Bell, R. D. et al. Pericytes control key neurovascular functions and neuronal phenotype in the adult brain and during brain aging. *Neuron* **68**, 409–427 (2010).
- Armulik, A. et al. Pericytes regulate the blood–brain barrier. *Nature* **468**, 557–561 (2010).
- Sullivan, P. M. et al. Reduced levels of human apoE4 protein in an animal model of cognitive impairment. *Neurobiol. Aging* **32**, 791–801 (2011).
- Satoh, K. et al. Cyclophilin A enhances vascular oxidative stress and the development of angiotensin II-induced aortic aneurysms. *Nature Med.* **15**, 649–656 (2009).
- Jin, Z. G. et al. Cyclophilin A is a proinflammatory cytokine that activates endothelial cells. *Arterioscler. Thromb. Vasc. Biol.* **24**, 1186–1191 (2004).
- Handschumacher, R. E., Harding, M. W., Rice, J., Drugge, R. J. & Speicher, D. W. Cyclophilin: a specific cytosolic binding protein for cyclosporin A. *Science* **226**, 544–547 (1984).
- Begley, D. J. et al. Permeability of the blood–brain barrier to the immunosuppressive cyclic peptide cyclosporin A. *J. Neurochem.* **55**, 1222–1230 (1990).
- Grammas, P. Neurovascular dysfunction, inflammation and endothelial activation: implications for the pathogenesis of Alzheimer's disease. *J. Neuroinflammation* **8**, 26 (2011).
- Paul, J., Strickland, S. & Melchor, J. P. Fibrin deposition accelerates neurovascular damage and neuroinflammation in mouse models of Alzheimer's disease. *J. Exp. Med.* **204**, 1999–2008 (2007).
- Zhong, Z. et al. Activated protein C therapy slows ALS-like disease in mice by transcriptionally inhibiting SOD1 in motor neurons and microglia cells. *J. Clin. Invest.* **119**, 3437–3449 (2009).
- Candelario-Jalil, E. et al. Matrix metalloproteinases are associated with increased blood–brain barrier opening in vascular cognitive impairment. *Stroke* **42**, 1345–1350 (2011).
- Garcia-Alloza, M. et al. Matrix metalloproteinase inhibition reduces oxidative stress associated with cerebral amyloid angiopathy *in vivo* in transgenic mice. *J. Neurochem.* **109**, 1636–1647 (2009).
- Jaeger, L. B. et al. Testing the neurovascular hypothesis of Alzheimer's disease: LRP-1 antisense reduces blood–brain barrier clearance, increases brain levels of amyloid-β protein, and impairs cognition. *J. Alzheimers Dis.* **17**, 553–570 (2009).
- Yang, Y., Lu, N., Zhou, J., Chen, Z. N. & Zhu, P. Cyclophilin A up-regulates MMP-9 expression and adhesion of monocytes/macrophages via CD147 signalling pathway in rheumatoid arthritis. *Rheumatology* **47**, 1299–1310 (2008).
- Deane, R. et al. apoE isoform-specific disruption of amyloid β peptide clearance from mouse brain. *J. Clin. Invest.* **118**, 4002–4013 (2008).
- Masliah, E. et al. Neurodegeneration and cognitive impairment in apoE-deficient mice is ameliorated by infusion of recombinant apoE. *Brain Res.* **751**, 307–314 (1997).
- Zipser, B. D. et al. Microvascular injury and blood–brain barrier leakage in Alzheimer's disease. *Neurobiol. Aging* **28**, 977–986 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank the National Institute of Health for grants R37NS34467 (B.V.Z.); R37AG23084 (B.V.Z.); RO1AG039452 (B.V.Z.); and R37AG13956 (D.M.H.) used to support this study.

Author Contributions R.D.B. designed and performed experiments, analysed data and contributed to writing the paper. E.A.W. designed and performed experiments. I.S. performed experiments. A.P.S. performed CBF experiments. R.D. designed experiments and analysed data. Z.W. gathered pilot data. D.M.H. provided guidance and edited the paper. C.B. designed experiments and edited the paper. A.A. performed pilot cadaverine studies. J.S. generated pilot data. B.C.B. provided guidance and edited the paper. B.V.Z. designed experiments, analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.V.Z. (zlokovic@usc.edu).

Glycolytic oligodendrocytes maintain myelin and long-term axonal integrity

Ursula Fünfschilling^{1*}, Lotti M. Supplie^{1*}, Don Mahad^{2*†}, Susann Boretius^{3*}, Aiman S. Saab^{1†}, Julia Edgar¹, Bastian G. Brinkmann^{1†}, Celia M. Kassmann¹, Iva D. Tzvetanova¹, Wiebke Möbius^{1†}, Francisca Diaz⁴, Dies Meijer⁵, Ueli Suter⁶, Bernd Hamprecht¹, Michael W. Sereda^{1,7}, Carlos T. Moraes⁴, Jens Frahm³, Sandra Goebbels¹ & Klaus-Armin Nave¹

Oligodendrocytes, the myelin-forming glial cells of the central nervous system, maintain long-term axonal integrity^{1–3}. However, the underlying support mechanisms are not understood⁴. Here we identify a metabolic component of axon–glia interactions by generating conditional *Cox10* (protoheme IX farnesyltransferase) mutant mice, in which oligodendrocytes and Schwann cells fail to assemble stable mitochondrial cytochrome *c* oxidase (COX, also known as mitochondrial complex IV). In the peripheral nervous system, *Cox10* conditional mutants exhibit severe neuropathy with dysmyelination, abnormal Remak bundles, muscle atrophy and paralysis. Notably, perturbing mitochondrial respiration did not cause glial cell death. In the adult central nervous system, we found no signs of demyelination, axonal degeneration or secondary inflammation. Unlike cultured oligodendrocytes, which are sensitive to COX inhibitors⁵, post-myelination oligodendrocytes survive well in the absence of COX activity. More importantly, by *in vivo* magnetic resonance spectroscopy, brain lactate concentrations in mutants were increased compared with controls, but were detectable only in mice exposed to volatile anaesthetics. This indicates that aerobic glycolysis products derived from oligodendrocytes are rapidly metabolized within white matter tracts. Because myelinated axons can use lactate when energy-deprived⁶, our findings suggest a model in which axon–glia metabolic coupling serves a physiological function.

The nuclear *Cox10* gene encodes a haem A farnesyl transferase, essential for the assembly of COX, the terminal complex of the electron transport chain in mitochondria^{7,8}. In *COX10* mutants, COX is unstable and rapidly degraded. *Cox10*^{flox/flox} mice have been used to generate models of severe mitochondrial disease in muscle, brain and liver^{7–10}. We crossed the *Cox10*^{flox/flox} line⁷ with *Cnp1*^{Cre/+} mice² that express Cre recombinase in Schwann cells and oligodendrocytes (with Cre expression beginning at the respective precursor stages in these cells)^{11,12}. In the absence of functional COX, glial cells should fail to fully metabolize glucose, and should generate ATP mostly by glycolysis and produce lactate. Thus, demyelination and glial cell death would be key indicators for metabolic demands of glial cells, toxic effects of lactic acidosis, and metabolic coupling between neuronal and glial compartments in myelinated fibre tracts.

Crosses of *Cox10*^{flox/+}**Cnp1*^{Cre/+} and *Cox10*^{flox/flox}**Cnp1*^{+/-} mice led to fewer mutant pups (*Cox10*^{flox/flox}**Cnp1*^{Cre/+}) than expected (11% compared with 25% expected, *n* = 466), suggesting prenatal death of some mice, probably owing to the known *Cnp1*^{Cre/+} expression in a subset of neural precursors (see later). We therefore generated other mutants by selectively targeting Schwann cells (using *Dhh-Cre* mice) and mature oligodendrocytes (using tamoxifen-inducible

Plp1-CreERT2 mice), and found no evidence of embryonic lethality. In the following, data are from *Cnp1*^{Cre/+}**Cox10*^{flox/flox} mutants (and *Cnp1*^{Cre/+}**Cox10*^{flox/+} controls), unless otherwise stated.

Quantitative polymerase chain reaction (qPCR) of genomic DNA from neural tissue (at postnatal day (P) 21) confirmed the loss of exon 6 and thus a functional *Cox10* gene (Fig. 1a, b). For the sciatic nerve we calculated that 67 ± 2% (mean ± s.e.m., *n* = 4) of all cells were recombinant (assuming that both alleles recombine). Thus, most Schwann cells in peripheral nerves were successfully targeted. In the optic nerve and cervical spinal cord, we calculated that about 50% of cells were recombined at the *Cox10* locus (Fig. 1b), the estimated percentage of oligodendrocytes¹³.

Elimination of *Cox10* is not immediately followed by the loss of respiration, but predicts the rapid functional ‘ageing’ of mitochondria, as the mitochondrial half-life is about 3 weeks in the brain^{14,15}. Thus, postnatal myelination by oligodendrocytes will involve intact mitochondria (Fig. 1c). By contrast, recombination of *Cox10* in proliferating Schwann cells causes a more rapid ‘loss’ of intact mitochondria by dilution in the progeny (Fig. 1c, right). Indeed, the absence of COX from many Schwann cells was obvious at P21 (Fig. 1d). For comparison, immunostaining of axonal mitochondria was indistinguishable between mutants and controls. We also noted occasional spinal motor neurons lacking COX (not shown), a plausible contributor to embryonic lethality.

A severe neuropathy phenotype proved that gene targeting was efficient and disruptive for mitochondrial function. Both *Cnp1*^{Cre/+}**Cox10*^{flox/flox} and *Dhh-Cre***Cox10*^{flox/flox} mice exhibited a developmental defect, with reduced weight, tremors and hindlimb weakness, although mutants appeared otherwise healthy and well groomed. The disease progressed to hindlimb paralysis and muscle atrophy (Supplementary Fig. 1a), followed by forelimb paralysis at about 9 months, when mutants had to be euthanized.

Already at P21 attempts to compare sciatic nerve conduction velocities between controls (15.6 ± 0.7 m s⁻¹; mean ± s.e.m.; *n* = 5) and mutant mice (*n* = 5) failed, indicating conduction blocks (Fig. 2a). Only distal stimulation elicited a compound muscle action potential in mutant mice, but with 50% reduced amplitude (Supplementary Fig. 1b), indicating a combination of glial and axonal perturbations. The same functional blocks were seen in *Dhh-Cre***Cox10*^{flox/flox} mice (not shown), confirming the specificity for Schwann cells.

By morphology and TdT-mediated dUTP nick end labelling (TUNEL) staining, we found no evidence for Schwann cell death as a possible cause of hypomyelination (not shown). In fact, the number of cell nuclei was increased (Fig. 2b). Compared with controls, mutant

¹Max Planck Institute of Experimental Medicine, Department of Neurogenetics, Hermann-Rein-Strasse 3, D-37075 Göttingen, Germany. ²Mitochondrial Research Group, Institute for Ageing and Health, The Medical School, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK. ³Biomedizinische NMR Forschungs GmbH am Max Planck Institute of Biophysical Chemistry, D-37070 Göttingen, Germany. ⁴University of Miami, Miller School of Medicine, Department of Neurology and Cell Biology & Anatomy, 1095 NW 14th Terrace, Miami, Florida 33136, USA. ⁵Department of Cell Biology, Erasmus MC, 3000 CA Rotterdam, Netherlands. ⁶Institute of Cell Biology, Department of Biology, ETH Hönggerberg, Schafmattstrasse 18, 8093 Zürich, Switzerland. ⁷Department of Clinical Neurophysiology, University of Göttingen (UMG), Robert-Koch-Strasse 40, D-37075 Göttingen, Germany. [†]Present addresses: Center for Neuroregeneration, University of Edinburgh, 49 Little France Crescent, Edinburgh EH16 4SB, UK (D.M.); Department of Molecular Physiology, Institute of Physiology, University of Saarland, D-66421 Homburg, Germany (A.S.S.); Department of Neurosurgery, Charité-University Medicine Berlin, Charitéplatz 1, D-10117 Berlin, Germany (B.G.B.); Department of Diagnostic Radiology, Am Botanischen Garten 14, University of Kiel, D-24118 Kiel, Germany.

*These authors contributed equally to this work.

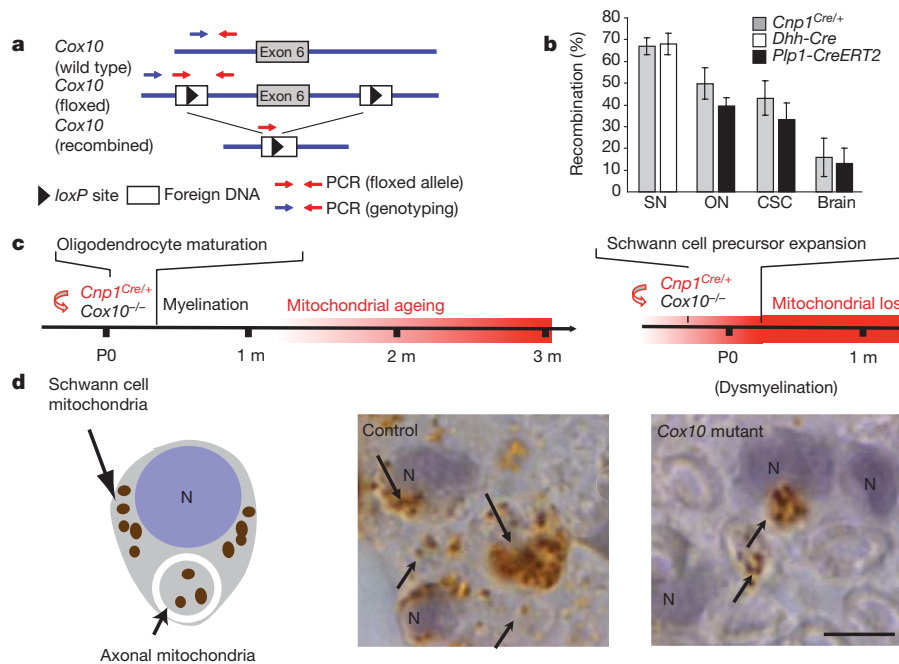


Figure 1 | Genetic targeting of the mitochondrial COX complex in myelinating glial cells.

a, Conditional mutagenesis of the floxed *Cox10* gene, deleting exon 6 (not drawn to scale). Arrows indicate the location of primers for genotyping (blue/red) and qPCR (red/red), the latter amplifying only the floxed allele. **b**, Loss of the floxed *Cox10* gene in myelinating glial cells of different Cre-expressing mice (adult), confirmed by qPCR of genomic DNA. Highest fraction of myelinating glia are found in sciatic nerve (SN), followed by optic nerve (ON), cervical spinal cord (CSC) and total brain. Mean percentages \pm s.d. are shown; $n = 3$ –5. **c**, Left panel, schematic time line showing that *Cox10* genomic recombination in immature oligodendrocytes of newborn *Cnp1^{Cre/+}* mice does not interfere with postnatal myelination. Mutant oligodendrocytes myelinate

nerves were thinner (Fig. 2c), the absolute number of myelinated axons was reduced (Fig. 2d) and many medium-sized axons remained unmyelinated (Supplementary Fig. 1c, d). However, neither large nor small axons were significantly altered in number (Supplementary Fig. 2a), suggesting a primary dysmyelination at P21. By immunostaining for the activated monocyte/macrophage marker MAC3, reactive macrophages were found to not yet be a feature (not shown), confirming the developmental defect. At later stages, the number of myelinated axons was reduced (Fig. 2d), and in 2-month-old mice fibre degeneration was already associated with secondary inflammation (Supplementary Fig. 2b). Notably, even at 9 months there was no major loss of Schwann cells, with 193 ± 25 per sciatic nerve section in mutants and 239 ± 17 in controls (mean \pm s.e.m., $n = 4$).

Where myelin was present, g-ratio analysis (the ratio between the diameter of the inner axon and the total outer diameter) showed a decrease in myelin thickness (0.589 ± 0.009 in controls; 0.612 ± 0.012 in mutants; mean \pm s.d., $n = 3$). By electron microscopy, many axons that appeared correctly sorted remained unmyelinated, despite a calibre $>1.0 \mu$ m (Fig. 2e–g). Remak cells failed to engulf single axons with cellular processes (Fig. 2f). Thus, the axonal sorting defect in mice lacking the transcription factor TFAM1 (ref. 16) is probably the consequence of perturbed COX activity in Schwann cells.

Mitochondria were visibly enlarged in mutant Schwann cells (Fig. 2h, i), but not in associated axons (Supplementary Fig. 2c), reflecting a loss of the proton gradient and a failure of H^+/K^+ antiporters¹⁷. Compacted myelin, when present, had a normal ultrastructure and was stable (Fig. 2h, i).

An intriguing aspect of altered energy metabolism in peripheral nerves was the increase of vascularization. Blood vessels covered a threefold larger area in cross-sections from mutant nerves (Supplementary Fig. 3a–c). Hypervascularization was associated with upregulation of *Vegfa*

(first 3 weeks), using pre-existing mitochondria that will subsequently decline in respiratory function (marked ‘ageing’). Right panel, for comparison, *Cox10* recombination in proliferating Schwann cell precursors perturbs mitochondrial function before myelination, as intact mitochondria are ‘lost’ by dilution with mutant mitochondria. M, month. **d**, Right panels, myelinated axons and Schwann cells in motor roots of control and mutant animals at P21. The left panel shows a schematic representation. *Cox10* mutant Schwann cells (but not axons) lose unstable COX, as demonstrated by reduced immunostaining for the catalytic COX subunit I (brown) in paraffin sections. Schwann cell nuclei (N) are counterstained in pale blue. Upward arrow, COX staining of axonal mitochondria; downward arrow, COX in Schwann cells. Scale bar, 10 μ m.

messenger RNA, but not of *Hif1a* or glucose transporters (Supplementary Fig. 3d), and is possibly a response to enhanced glycolysis.

Myelinated tracts were normally developed in brains and spinal cords of mice at 2 months of age (Supplementary Fig. 4), which is best explained by sufficient respiration of oligodendroglial mitochondria during myelination (Fig. 1c). The fact that losing mitochondrial function requires several weeks was confirmed by inactivating *Cox10* in postmitotic neurons (at P5–P10) using *CamKII-Cre* mice. As previously shown¹⁰, these mice die prematurely with severe neurodegeneration, visualized by neuroinflammation and loss of hippocampal neurons at 4 months of age (Supplementary Fig. 5).

Surprisingly, even at 9 months of age when *Cnp1^{Cre/+}*Cox10^{fllox/fllox}* mice had to be killed because of peripheral neuropathy, we detected no signs of white matter pathology, demyelination or oligodendroglial pathology (Fig. 3a–f), although the absence of COX activity from oligodendrocyte lineage cells was obvious by sequential COX and succinate dehydrogenase (SDH, also known as mitochondrial complex II) histochemistry (Fig. 3g and Supplementary Fig. 6). Mature oligodendrocytes ($CC1^+$) instead showed an abnormal (SDH^+) mitochondrial expansion, but were COX^- (Fig. 3h). To extend the observational time window, we induced recombination by tamoxifen in *Plp1-CreERT2**Cox10^{fllox/fllox}** 1-month-old mice. Histological analysis at age 14 months (the latest time-point tested) also failed to show any signs of demyelination or neurodegeneration (Supplementary Fig. 7a–c). Furthermore, magnetic resonance imaging (MRI) showed that there was no difference in ventricular volume between mutants ($8.1 \pm 1.1 \text{ mm}^3$ (mean \pm s.d.)) and controls ($8.4 \pm 1.4 \text{ mm}^3$) at 6–7 months of age (Supplementary Fig. 7d, e).

Theoretically, white matter integrity could be preserved when mutant oligodendrocytes are replaced by newly generated oligodendrocytes that remyelinate. These cells might again survive for several weeks after recombination of *Cox10*. We think this is unlikely for

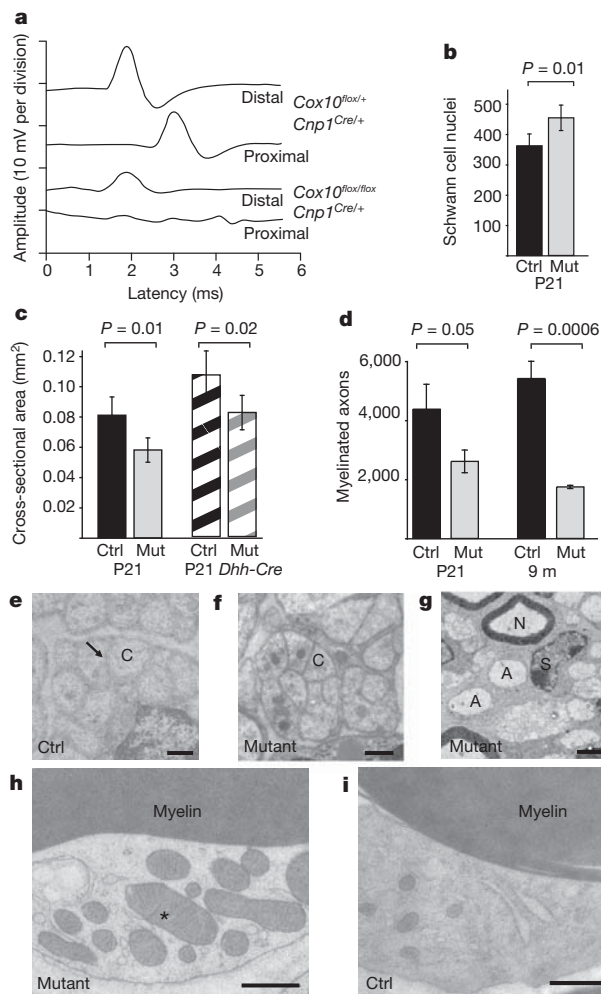


Figure 2 | Peripheral neuropathy caused by *Cox10* mutant Schwann cells. **a**, Amplitudes of compound muscle action potentials were recorded in controls and mutant mice at P21 after proximal or distal stimulation of sciatic nerves, but were barely detectable in the mutants. **b–d**, The morphology of sciatic nerves was assessed in semithin sections, and the number of Schwann cell nuclei was counted (**b**), the cross-sectional area of the nerve with fascicles of axons and Schwann cells was determined (**c**), and the absolute number of myelinated axons counted (**d**). This showed a progressive loss of axons at older age, but not of Schwann cells. All data are mean \pm s.d. Ctrl, control; mut, mutant. **e–i**, Electron microscopy showed differences between mutant and control sciatic nerves. **e, f**, In Remak bundles, C-fibre axons (C) are normally sorted by Schwann cell processes (e, arrow) but not in mutants (**f**). **g**, Mutant Schwann cells that fail to myelinate survive well. Note numerous unmyelinated medium-calibre axons (A) next to normally myelinated axons (N) and intact Schwann cell nuclei (S). **h, i**, At higher magnification, myelin appeared morphologically normal, but mutant mitochondria (asterisk) were clearly enlarged. Scale bars, 500 nm (**e, f, h, i**), 2 μ m (**g**).

several reasons. By electron microscopy and *g*-ratio analysis, central nervous system (CNS) myelin thickness in adult mutants was in the normal range (Fig. 3c, d and Supplementary Fig. 8), unlike remyelinated axons that are hypomyelinated¹⁸. Moreover, we failed to observe apoptotic cells in the white matter tracts of adult *Cox10* mutants (Supplementary Fig. 9a), and the administration of BrdU did not label more oligodendrocyte precursor cells in mutants than in age-matched controls (Supplementary Fig. 9b). Furthermore, staining for astrocytes, microglia and T cells (Supplementary Fig. 10) gave no indication of low-grade inflammation, a very sensitive response to degenerative processes in white matter tracts. We conclude that once myelination has occurred, reduced mitochondrial functions do not perturb oligodendrocyte survival, myelin maintenance or axonal integrity.

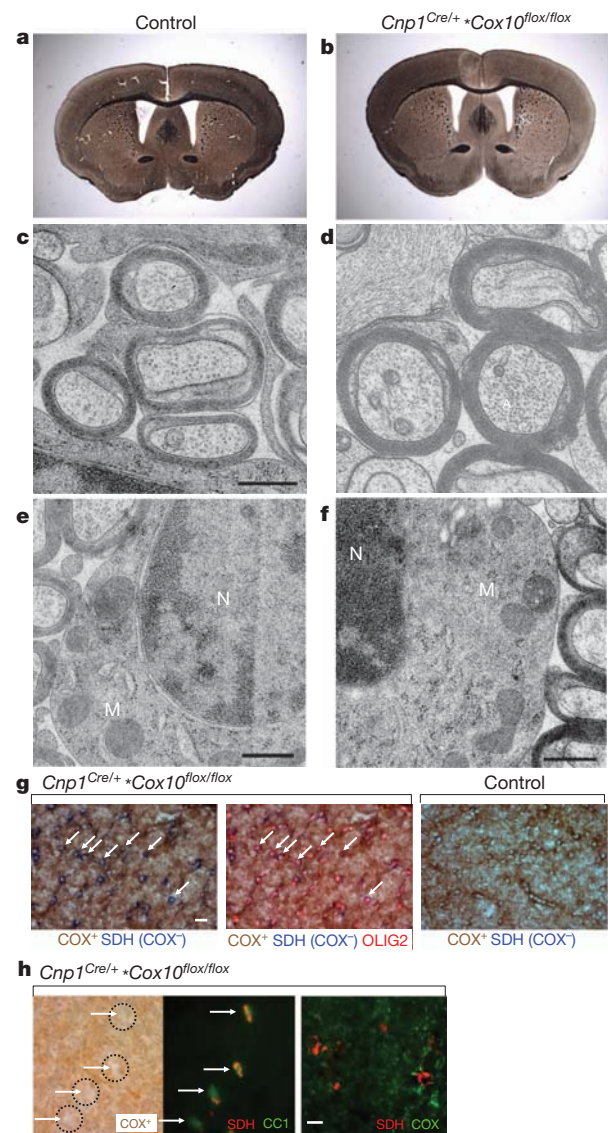
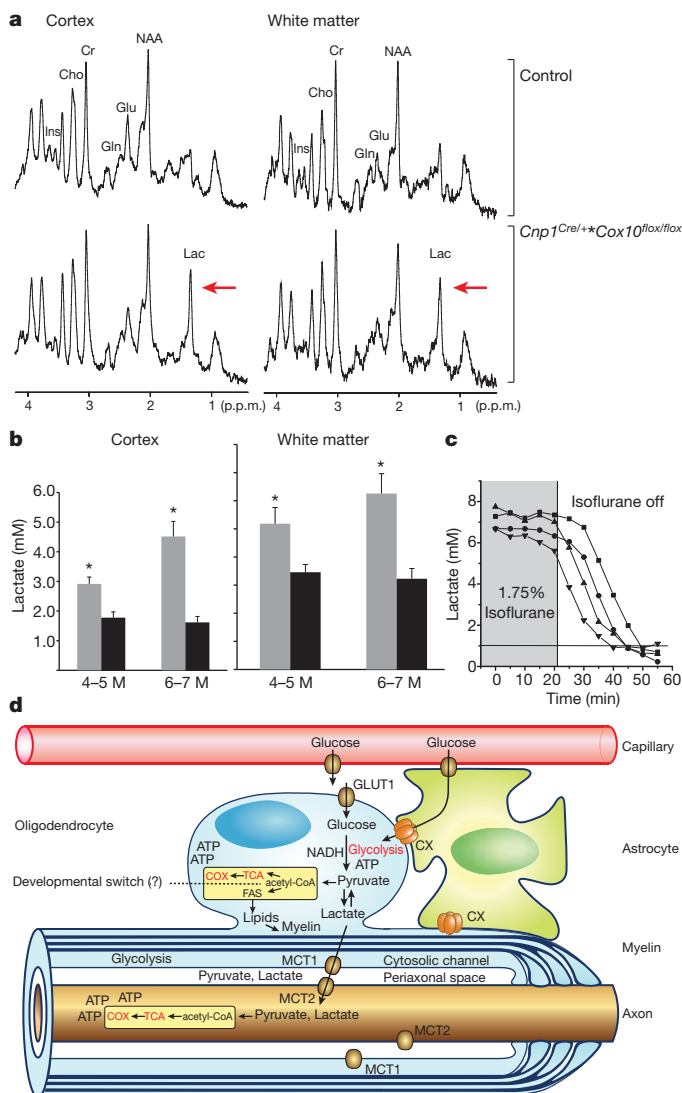


Figure 3 | Oligodendroglial survival, myelin preservation and white matter integrity in *Cnp1^{Cre/+} * Cox10^{flx/flx}* mice. **a, b**, By Gallyas' silver impregnation of myelin at 9 months of age, the corpus callosum and other white matter tracts appear normally developed and stable in mutant mice. **c–f**, Electron microscopy of high-pressure frozen optic nerve shows intact myelination of CNS axons (**c, d**), and healthy oligodendroglia nuclei (**e, f**). A, axon; N, nucleus; M, mitochondria. Scale bars, 0.5 μ m. **g**, Serial COX and SDH histochemistry. Left, in the normal appearing white matter of 9-month-old *Cnp1^{Cre/+} * Cox10^{flx/flx}* mutants (corpus callosum), mitochondrial COX activity yields a brown precipitate (axons and astrocytes, for example). By serial COX and SDH histochemistry, only *Cox10*-mutant COX[−] cells are visibly stained for SDH (blue precipitate, white arrows). Middle, about half (48.2 \pm 6.5%) of the OLIG2-positive oligodendrocyte lineage cells in the mutant corpus callosum (red nuclei in overlay), that is, mostly mature oligodendrocytes that express CNP1^{Cre}, are lacking COX activity (blue). Right, there are no COX[−] cells in the corpus callosum of age-matched controls. Scale bar, 20 μ m. **h**, Left, when combining only COX histochemistry with a marker for mature cells (CC1), mutant oligodendrocytes (corpus callosum; 9 months of age) appear white (arrows), intermingled with COX⁺ neighbouring cells and compartments. Middle, mature oligodendrocytes that are CC1⁺ (green) are filled with abnormally large numbers of SDH⁺ mitochondria (red), and appear in orange in the overlay. Note the near absence of SDH signals from neighbouring cells and compartments. Right, in a separate overlay, the SDH⁺ mitochondria (red) are virtually devoid of COX immunostaining (green) in mutant oligodendrocytes. Scale bar, 10 μ m.

We proposed that oligodendrocytes survive by enhanced glycolysis, and predicted to see *in vivo* an increase in the brain lactate concentration using localized proton magnetic resonance spectroscopy (MRS). As expected, *Cox10* mutants showed a more strongly increased lactate resonance (compared with controls) in both cortex and white matter. Notably, accumulated brain lactate was detectable only under isoflurane anaesthesia (Fig. 4a). When quantified in adult mutants (Fig. 4b), we determined >4 mM lactate in the cortex and >6 mM in the corpus callosum (compared with <2 mM lactate in the cortex and <4 mM in the white matter of wild-type mice). At the end of anaesthesia, these lactate concentrations fell within minutes to very low levels (Fig. 4c). The latter finding is compatible with a model in which oligodendroglial release of lactate — the necessary by-product of aerobic glycolysis — is followed by its rapid use in other cellular compartments. It is demonstrated *in vitro* that glial lactate is efficiently metabolized by myelinated axons (Supplementary Fig. 11). *In vivo*, lactate levels never reached MRS detectability under physiological conditions, unless challenged by isoflurane anaesthesia. Of note, the levels of *N*-acetylaspartate (a marker of viable neurons and axons), choline-containing compounds (indicators of membrane turnover) and myo-inositol (an osmolyte and glial membrane constituent) further support our conclusion that neurodegeneration and abnormal oligodendrocyte turnover are not a feature (Supplementary Fig. 12).

COX10-deficient cells in culture are practically respiration-deficient¹⁹. Because we combined well-established lines of floxed mice and Cre transgenic tools, the reported phenotype is probably caused by



the loss of mitochondrial ATP generation. In the periphery, the mitotic expansion of (mutant) Schwann cell precursors rapidly dilutes functional mitochondria, and hypomyelination demonstrates that myelination is an energy-consuming process. We also note that many patients with mitochondrial disorders have signs of neuropathy with hypomyelination. By contrast, unperturbed CNS myelination is best explained by sufficient residual function of mitochondria in mature oligodendrocytes after recombination (Fig. 1c). We note that germline mutations of human *COX10* and other COX assembly factors, although ultimately lethal, have been associated with CNS dysmyelination (leukodystrophy)⁸.

Most intriguing is the mild post-myelination CNS phenotype of adult *Cox10* mutant mice, in which oligodendrocytes maintain myelin and their ensheathed axons. That mature oligodendrocytes apparently survive by aerobic glycolysis similar to tumour cells (known as the Warburg effect) was unexpected. In fact, oligodendrocytes *in vitro* are sensitive to the COX inhibitor azide⁵. However, these cells correspond to ‘pre-myelinating’ oligodendrocytes *in vivo*, and it seems that metabolic properties of oligodendrocyte lineage cells change during maturation. After myelination, oligodendrocytes may tolerate a partly glycolytic metabolism that we have experimentally augmented in *Cox10* mutants. We note that mature wild-type oligodendrocytes have lower COX activity than oligodendrocyte precursor cells (not shown), and that white matter has relatively high glycolytic activity²⁰. Also, in our control mice absolute lactate levels (as a surrogate marker for aerobic glycolysis) were higher in white than grey matter (Fig. 4b).

Enhanced glycolysis in oligodendrocytes increases the production of pyruvate, lactate and acetyl-coenzyme A (CoA), the latter promoting fatty acid synthesis during myelination. Because energy-deprived myelinated axons can rapidly use pyruvate and lactate as an alternative energy source^{6,21} (Supplementary Fig. 11), we suggest a model in which an increased rate of oligodendroglial glycolysis can supply glycolysis products to support axonal energy needs (Fig. 4d).

Our observation that lactate levels in mutant brains are visibly increased compared with controls, but only under isoflurane anaesthesia, suggests that lactate is not cleared by drainage, but is locally metabolized. Isoflurane anaesthesia does not decrease CNS blood flow. The theoretical possibility that increased lactate peaks are caused by mutant oligodendrocytes that fail to metabolize neuronal lactate seems remote (note the similar lactate increases in grey matter, where neurons are associated with respiration-competent astrocytes²²). Increased

Figure 4 | Rapid use of lactate shown by proton MRS. **a**, Localized proton magnetic resonance spectra of the cortex (left) and corpus callosum (right) from *Cnp1^{Cre/+};Cox10^{lox/lox}* mutants and controls (6–7 months of age). Cr, total creatine; Ins, myo-inositol; p.p.m., parts per million; NAA, *N*-acetylaspartate. Red arrows denote lactate (Lac). **b**, Lactate levels in the cortex (left) and white matter (right) are increased in mutant mice (grey bars) compared with controls (black bars) under isoflurane anaesthesia. Data are mean \pm s.e.m.; $n = 6$ –7 per genotype. Note that under isoflurane anaesthesia the control mice have higher lactate levels in white matter than in grey matter. M, months. **c**, Increased brain lactate levels drop to undetectable levels in less than 60 min at the end of isoflurane anaesthesia. This suggests that lactate (produced by oligodendrocytes) is rapidly metabolized by other cellular compartments in the white matter tracts of awake mice. **d**, Hypothetical model of metabolic coupling between oligodendrocytes and myelinated axons. Oligodendrocytes import glucose through GLUT1 (and possibly via astrocytes and gap junctions; CX, connexin) for glycolysis. Pyruvate is metabolised in mitochondria (yellow) for ATP generation (TCA, tricarboxylic acid cycle). With the onset of myelination (‘developmental switch’), glucose also serves the synthesis of fatty acid (FAS) and myelin lipids from acetyl-CoA. In post-myelination oligodendrocytes, glycolysis can yield sufficient ATP to support oligodendrocyte survival. Glycolysis products are used by myelinated axons when energy levels are low^{6,21}. Lactate (or pyruvate when NADH is oxidized in oligodendroglial mitochondria) can be directly transferred via monocarboxylic acid transporters (MCT1, MCT2), which reside in internodal myelin²⁴ and the axonal compartment, such that lactate is rapidly cleared *in vivo* (in c). Note that myelinated axons, largely shielded from the extracellular milieu⁴, are separated by a thin periaxonal space from the oligodendroglial cytoplasm filling the inner loops of myelin (‘cytosolic channel’) and paranodal loops.

lactate concentrations probably originate from *Cox10* mutant oligodendrocytes that rely on increased glycolysis for survival. Even under normal conditions, lactate exported from oligodendrocytes would be rapidly metabolized and therefore not be detected by MRS.

By *in situ* hybridization²³ and immunoelectron microscopy²⁴, the monocarboxylate transporter MCT1 (also known as SLC16A1) is expressed in white matter, and is even present in myelin²⁴, whereas MCT2 localizes to myelinated axons. This strongly suggests that lactate generated in oligodendrocytes is released and reaches the axonal compartment (Fig. 4d).

We assume that oligodendrocytes can oxidize glycolysis-derived NADH⁺ when mitochondria are intact (using a NADH-glycerolphosphate shuttle that is oligodendrocyte-specific²⁵). Thus, normal oligodendrocytes may not even reduce pyruvate, which passes equally well through MCT1 and MCT2. Such a model of oligodendrocytes, yielding glycolysis products to axons when energy deprived⁴, does not contradict the presumed role of lactate in the support of myelination during development²⁴, that is, when the import of metabolites is rate-limiting for lipid synthesis²⁶, specifically for oligodendrocytes in culture²⁴.

Our model of axon–oligodendrocyte metabolic coupling is reminiscent of the astroglial ‘lactate shuttle’, which is thought to support the energy metabolism of glutamatergic synapses in the cortex^{27,28}. For oligodendrocyte lineage cells, the inferred developmental ‘switch’ of metabolism even allows them to survive by aerobic glycolysis. The underlying mechanisms deserve further attention, as they may be relevant for the survival of glioma cells and the Warburg effect.

METHODS SUMMARY

Animal experiments were in compliance with regulations of the State of Lower Saxony, Germany. Mice were maintained on a C57/Bl6 genetic background and genotyped as described (for references, see Methods). For *in vivo* electrophysiology of anesthetized mice, recording electrodes were placed at the intrinsic foot muscle, distal stimulation electrodes at the ankle, and proximal electrodes at the sciatic notch. Tamoxifen treatment was for 5 days, daily. qPCR was performed in triplicates. For serial COX and SDH histochemistry, cryostat sections were air-dried before incubation with cytochrome c, diaminobenzidine tetrahydrochloride and catalase, followed by the incubation with sodium succinate, phenazine methosulphate, sodium azide and nitroblue tetrazolium. Gallyas’ silver impregnation has been described³. Immunohistochemistry was on paraffin-embedded material. Tissues for cryosections were snap-frozen or immersion-fixed, infiltrated with sucrose and frozen in Shandon Cryomatrix. TUNEL assays were performed using DeadEndTM Colorimetric kit (Promega). For labelling mitotic cells, BrdU was injected intraperitoneally at 4 months of age, 5 days a week for 4 weeks. For primer sequences and antibodies see Methods. For electron microscopy, mice were perfusion-fixed with 2.5% glutaraldehyde and 4% paraformaldehyde. Dissected nerves were postfixed overnight and embedded in Epon. Semithin sections were stained with methylene blue and AzurII. Ultrathin sections were contrasted with 1% uranyl acetate and lead citrate. Optic nerves were also cryofixed by high-pressure freezing and freeze substituted. For MRI and MRS, mice were initially anaesthetized with 5% isoflurane, intubated and kept under anaesthesia with 1.75% isoflurane in ambient air. *In vivo* localized proton MRS in different regions of the brain was performed at 9.4 T. T2-weighted MRI ensured a proper position of volumes-of-interest. Metabolite quantification involved spectral evaluation and calibration with brain water concentration.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 June 2011; accepted 2 March 2012.

Published online 29 April 2012.

- Griffiths, I. *et al.* Axonal swellings and degeneration in mice lacking the major proteolipid of myelin. *Science* **280**, 1610–1613 (1998).
- Lappe-Siefke, C. *et al.* Disruption of *Cnp1* uncouples oligodendroglial functions in axonal support and myelination. *Nature Genet.* **33**, 366–374 (2003).
- Kassmann, C. M. *et al.* Axonal loss and neuroinflammation caused by peroxisome-deficient oligodendrocytes. *Nature Genet.* **39**, 969–976 (2007).
- Nave, K. A. Myelination and the trophic support of long axons. *Nature Rev. Neurosci.* **11**, 275–283 (2010).
- Ziabreva, I. *et al.* Injury and differentiation following inhibition of mitochondrial respiratory chain complex IV in rat oligodendrocytes. *Glia* **58**, 1827–1837 (2010).
- Tekkök, S. B., Brown, A. M., Westenbroek, R., Pellerin, L. & Ransom, B. R. Transfer of glycogen-derived lactate from astrocytes to axons via specific monocarboxylate transporters supports mouse optic nerve activity. *J. Neurosci. Res.* **81**, 644–652 (2005).
- Diaz, F., Thomas, C. K., Garcia, S., Hernandez, D. & Moraes, C. T. Mice lacking *COX10* in skeletal muscle recapitulate the phenotype of progressive mitochondrial myopathies associated with cytochrome c oxidase deficiency. *Hum. Mol. Genet.* **14**, 2737–2748 (2005).
- Antonicka, H. *et al.* Mutations in *COX10* result in a defect in mitochondrial heme A biosynthesis and account for multiple, early-onset clinical phenotypes associated with isolated COX deficiency. *Hum. Mol. Genet.* **12**, 2693–2702 (2003).
- Diaz, F. *et al.* Pathophysiology and fate of hepatocytes in a mouse model of mitochondrial hepatopathies. *Gut* **57**, 232–242 (2008).
- Fukui, H., Diaz, F., Garcia, S. & Moraes, C. T. Cytochrome c oxidase deficiency in neurons decreases both oxidative stress and amyloid formation in a mouse model of Alzheimer’s disease. *Proc. Natl Acad. Sci. USA* **104**, 14163–14168 (2007).
- Goebbels, S. *et al.* Elevated phosphatidylinositol 3,4,5-trisphosphate in glia triggers cell-autonomous membrane wrapping and myelination. *J. Neurosci.* **30**, 8953–8964 (2010).
- Genoud, S. *et al.* Notch1 control of oligodendrocyte differentiation in the spinal cord. *J. Cell Biol.* **158**, 709–718 (2002).
- Miller, R. H., David, S., Patel, R., Abney, E. R. & Raff, M. C. A quantitative immunohistochemical study of macroglial cell development in the rat optic nerve: *in vivo* evidence for two distinct astrocyte lineages. *Dev. Biol.* **111**, 35–41 (1985).
- Beattie, D. S., Basford, R. E. & Koritz, S. B. The turnover of the protein components of mitochondria from rat liver, kidney, and brain. *J. Biol. Chem.* **242**, 4584–4586 (1967).
- Menzies, R. A. & Gold, P. H. The turnover of mitochondria in a variety of tissues of young adult and aged rats. *J. Biol. Chem.* **246**, 2425–2429 (1971).
- Viader, A. *et al.* Schwann cell mitochondrial metabolism supports long-term axonal survival and peripheral nerve function. *J. Neurosci.* **31**, 10128–10140 (2011).
- Garlid, K. D. & Paucek, P. Mitochondrial potassium transport: the K⁺ cycle. *Biochim. Biophys. Acta* **1606**, 23–41 (2003).
- Dubois-Dalcq, M., French-Constant, C. & Franklin, R. J. Enhancing central nervous system remyelination in multiple sclerosis. *Neuron* **48**, 9–12 (2005).
- Diaz, F., Fukui, H., Garcia, S. & Moraes, C. T. Cytochrome c oxidase is required for the assembly/stability of respiratory complex I in mouse fibroblasts. *Mol. Cell. Biol.* **26**, 4872–4881 (2006).
- Moreland, C., Henjum, S., Iversen, E. G., Skredde, K. K. & Hassel, B. Evidence for a higher glycolytic than oxidative metabolic activity in white matter of rat brain. *Neurochem. Int.* **50**, 703–709 (1990).
- Brown, A. M., Wender, R. & Ransom, B. R. Metabolic substrates other than glucose support axon function in central white matter. *J. Neurosci. Res.* **66**, 839–843 (2001).
- Gandhi, G. K., Cruz, N. F., Ball, K. K. & Dienes, G. A. Astrocytes are poised for lactate trafficking and release from activated brain and for supply of glucose to neurons. *J. Neurochem.* **111**, 522–536 (2009).
- Vannucci, S. J. & Simpson, I. A. Developmental switch in brain nutrient transporter expression in the rat. *Am. J. Physiol. Endocrinol. Metab.* **285**, E1127–E1134 (2003).
- Rinholm, J. E. *et al.* Regulation of oligodendrocyte development and myelination by glucose and lactate. *J. Neurosci.* **31**, 538–548 (2011).
- Leveille, P. J., McGinnis, J. F., Maxwell, D. S. & de Vellis, J. Immunocytochemical localization of glycerol-3-phosphate dehydrogenase in rat oligodendrocytes. *Brain Res.* **196**, 287–305 (1980).
- Jailil, M. A. *et al.* Reduced N-acetylaspartate levels in mice lacking aralar, a brain- and muscle-type mitochondrial aspartate-glutamate carrier. *J. Biol. Chem.* **280**, 31333–31339 (2005).
- Pellerin, L. & Magistretti, P. J. Glutamate uptake into astrocytes stimulates aerobic glycolysis: a mechanism coupling neuronal activity to glucose utilization. *Proc. Natl Acad. Sci. USA* **91**, 10625–10629 (1994).
- Suzuki, A. *et al.* Astrocyte-neuron lactate transport is required for long-term memory formation. *Cell* **144**, 810–823 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Stiles for the OLIG2 antibodies, A. Fahrenholz, U. Bode, T. Ruhwedel, and R. Tammer for technical support, and members of the Nave laboratory for discussions. We acknowledge grant support from the BMBF (Leukonet), DFG (CMPB), EU-FP7 programs (NGIDD, Leukotreat) and Oliver’s Army. U.S. is supported by the Swiss National Science Foundation and the National Center ‘Neural Plasticity and Repair’. U.F. was supported by fellowships from the EU-FP7 (Marie-Curie), the Swiss National Science Foundation (PA00A-117479/1) and the European Leukodystrophy Association. K.-A.N. holds an ERC Advanced Grant.

Author Contributions U.F., L.M.S., C.M.K. and I.D.T. performed mouse breeding experiments, histology and light microscopy; D.Ma. carried out immunohistochemistry; S.B. performed magnetic resonance imaging and spectroscopy; A.S.S. and J.E. carried out *ex vivo* experiments; B.G.B. and M.W.S. performed electrophysiology; W.M. performed electron microscopy; F.D. and C.T.M. provided floxed mice; D.Mi. and U.S. provided Cre-transgenic lines. B.H., J.F. and S.G. supervised parts of the work or contributed essential ideas. K.-A.N. designed experiments, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.-A.N. (nave@em.mpg.de).

METHODS

Animals. All animal experiments were carried out in compliance with animal policies of the State of Lower Saxony, Germany. Animals were maintained on a C57/Bl6 genetic background. *Cox10^{fllox/fllox}* mice⁷, *Cnpl^{Cre/+}* mice², *Dhh-Cre* mice²⁹, *Plp1-CreERT2* mice³⁰ and *CamKII-Cre³¹* mice were genotyped as described. Tamoxifen treatment³⁰ was for 5 days, daily.

qPCR of genomic DNA. qPCR was performed using 10 ng of genomic DNA in a 12.5- μ l assay using PowerSYBR Green PCR Master Mix (Applied Biosystems) on an Applied Biosystems 7500 Fast Real Time PCR system in triplicates. A standard curve was generated using genomic DNA from *Cox10^{fllox/fllox}*, *Cox10^{fllox/+}* and *Cox10^{+/+}* animals. The values were standardized to an independent genomic marker. A 167-base-pair *Cox10^{fllox}*-specific fragment was obtained with primers 5'-CGGGGATCAATTCGAGCTCGCC-3' and 5'-CACTGACGACGCGCCAGCATCTT-3'. The percentage of *Cox10* recombination was calculated by assuming that recombination affects both floxed alleles in Cre-expressing cells.

Quantitative PCR after reverse transcription. RNA was prepared in TRIzol (Invitrogen), followed by purification on a Qiagen RNeasy column. Complementary DNA was synthesized from 0.5 μ g of total RNA using SuperscriptIII (Invitrogen) with random nonamer primers. qPCRs were run with PowerSYBR Green PCR Master Mix (Applied Biosystems) on an Applied Biosystems 7500 Fast real-time PCR system in quadruplicates (outliers were removed). The following primers were used: *Gapdh*, 5'-CAATGAATACGGCTACAGCAAC-3' and 5'-TTACTCCTTGAGGCCATGT-3'; *Glut-1* (also known as *Slc2a1*), 5'-ATGGATCCAGCAGCAAG-3' and 5'-CCAGTGTTATAGCCGAACTGC-3'; *Glut-3* (*Slc2a3*), 5'-CTGGACCTCCAACCTTCTGG-3' and 5'-GCAGCGAAGATGATAAAACG-3'; *Hif1a*, 5'-CATGATGGCTCCCTTTTCA-3' and 5'-GTCACTGGTGTGCAATA-3'; *Vegfa*, 5'-TTACTGCTGTACCTCCACC-3' and 5'-ACAGGACGGCTTGAAGAT-3'; 18S RNA: 5'-AAATCAGTTATGGTTCCTTGGTC-3' and 5'-GCTCTAGAATTACCACAGTTATCCAA-3'.

Electrophysiology. Mice were anaesthetized with xylacin and ketamin intraperitoneally. Two recording electrodes were inserted into the intrinsic foot muscle. Distal stimulation electrodes were inserted at the ankle. The proximal stimulation electrodes were inserted at the sciatic notch. Compound muscle action potentials (CMAPs) were recorded with a Jaeger-Toennies Neuroscreen instrument. Nerve conduction velocities were calculated from the distance between proximal and distal stimulation electrodes and the latency difference between the CMAPs after successive proximal and distal stimulation. CMAP amplitudes were calculated peak to peak.

Optic nerve recordings and oxygen glucose deprivation. Optic nerves from 2-month-old wild-type mice were gently removed and directly transferred into an interface perfusion chamber (Haas Top, Harvard Apparatus) and superfused with artificial cerebrospinal fluid (ACSF) containing: 124 mM NaCl, 3 mM KCl, 2 mM CaCl₂, 2 mM MgSO₄, 1.25 mM NaH₂PO₄, 23 mM NaHCO₃ and 10 mM glucose. The perfusion chamber was continuously aerated by a humidified mixture of 95% O₂ and 5% CO₂, and experiments were performed at 37 °C. Custom-made suction electrodes, back-filled with ACSF, were used for stimulation and recording. Compound action potentials (CAP) were evoked every 30 s with a square-wave (50 μ s, 0.8 mA) constant current stimulus pulse (Stimulus Isolator 385, WPI). The recording electrode was connected to a Heka amplifier (EPC9), and the signal was amplified 500 times, filtered at 30 kHz, and acquired at 20–30 kHz. Before recording, optic nerves were equilibrated for at least 15 min in the chamber. After 1-h baseline recordings, optic nerves were exposed to oxygen glucose deprivation (OGD) for 1 h. OGD was induced by switching to glucose-free ACSF and a gas mixture containing 95% N₂ and 5% CO₂ in the recording chamber. To ensure that no oxygen was delivered by the superfusate, ACSF was always bubbled with 95% N₂, 5% CO₂. After OGD, control ACSF and O₂ were restored and CAPs were recorded for up to 2 h. To test the role of pyruvate and lactate transporters on energetic recovery, 200 μ M α -cyano-4-hydroxycinnamic acid (Sigma Aldrich) was applied 20 min before recovery (for 40 min). Optic nerve function was monitored over time, as the area under the CAP that was normalized to baseline. Data are presented as mean \pm s.e.m., with *n* indicating the number of optic nerves used in these experiments.

Serial COX and SDH histochemistry. Cryostat sections (8- μ m thickness) were air-dried for 30 min before the application of COX medium (500 μ M cytochrome c, 5 mM diaminobenzidine tetrahydrochloride and 20 mg ml⁻¹ catalase). After incubation in COX medium at 37 °C for 40 min, the sections were washed three times using PBS and SDH medium (130 mM sodium succinate, 200 mM phenazine methosulphate, 1 mM sodium azide, 1.5 mM nitroblue tetrazolium in 0.1 M phosphate buffer, pH 7.0) was applied for 30 min at 37 °C.

Histology and immunohistochemistry. Animals were anaesthetized with avertin and perfused intracardially with 15 ml of HBSS, followed by 50 ml of fixative (4% formalin in phosphate buffer). Dissected nerves were postfixed overnight in

fixative and embedded in paraffin. Gallyas' silver impregnation was performed as described². Tissues for cryosections were dissected fresh, and snap-frozen or immersion-fixed for 1 h in fixative, infiltrated with 20% sucrose in PBS, and frozen in Shandon Cryomatrix. Antibodies used were mouse-anti-COX subunit-I (MitoSciences), rat-anti-CD31 (BD Pharmingen), anti-GFAP (Chemicon), anti-MAC3 (BD Pharmingen) and anti-CD3 (Serotec). Images were processed in Adobe Photoshop to equalize the contrast, and the background around semithin sections was manually cleared.

For immunofluorescent labelling of OLIG2 after COX and SDH histochemistry, the sections were fixed in cold 4% paraformaldehyde solution for 30 min and antigen retrieval was performed using boiling EDTA buffer, pH 8, for 1 min. Sections were washed in TBS three times for 5 min. Non-specific binding was blocked using 1% normal goat serum (Sigma) diluted in TBS. Sections were then incubated with primary antibody (OLIG2, rabbit immunoglobulin-G (IgG); Chemicon) overnight at 4 °C and detected using rhodamine-conjugated anti-rabbit secondary antibody (Jackson ImmunoResearch). Sections were mounted using aqueous medium.

For double-labelling mitochondria in mature oligodendrocytes, we used COX-I (isotype IgG2a) and SDH (IgG1) antibodies from Mitosciences, and CC1 (IgG2b) antibodies from Calbiochem. Cryostat sections were air-dried for 30 min and fixed in cold 4% paraformaldehyde for 30 min. Gentle antigen retrieval was performed by immersing the section in boiling EDTA buffer, pH 7.4, for 30 s. After blocking in 1% normal goat serum, primary antibodies (1:200 for SDH 70 kDa and COX-I, and 1:100 for CC1) were applied for 90 min at room temperature. Indirect detection was with directly conjugated Alexia Fluor secondary antibodies (30 min at room temperature), and sections were mounted in Vector shield.

The TUNEL assay was performed using DeadEndTM Colorimetric TUNEL System kit following the manufacturer's protocol (Promega). For labelling mitotic cells, BrdU (100 μ g g⁻¹) was injected (intraperitoneally) at 4 months of age (for 4 weeks at 5 days a week). Brains were embedded in paraffin and sections (5 μ m) were stained with anti-BrdU (Chemicon) and anti-OLIG2 antibodies (a gift from from C. Stiles).

Semithin sections and electron microscopy. Animals were anaesthetized with avertin and perfused intracardially with 15 ml of HBSS, followed by 50 ml of fixative (2.5% glutaraldehyde, 4% paraformaldehyde in PBS). Dissected nerves were postfixed overnight in fixative and embedded in Epon. Semithin sections were stained with methylene blue and AzurII. Ultrathin sections were contrasted with 1% uranyl acetate and lead citrate. Electron microscopic pictures were taken on a LEO 912AB electron microscope (Zeiss) with an on-axis 2048 \times 2048 charge coupled device camera (Proscan). All cell counts and tracings were done using ImageJ (<http://rsb.info.nih.gov/ij/>). For semithin sections, entire cross-sections of sciatic nerves were scored. Optic nerves were freshly isolated and cryofixed by high-pressure freezing and freeze substituted. On electron microscopic pictures, all axons in three random fields (5,500 μ m² total) per nerve were traced. The g-ratios were calculated from circular areas equal to the measured area of the fibres.

MRI and MRS. MRI and MRS was performed as recently described³² on mutant and controls (*Cox10^{fllox/+} *Cnpl^{Cre/+}*) at 4–5 months (*n* = 6/6) and 6–7 months (*n* = 7/7). Mice were initially anaesthetized with 5% isoflurane, subsequently intubated and kept under anaesthesia with 1.75% isoflurane in ambient air. In four mice, further spectra were obtained without anaesthesia. In these mice, pancuronium (15 mg kg⁻¹) was administered 15 min before switching off the isoflurane supply to avoid movement artefacts. *In vivo* localized proton MRS (STEAM, TR/TE/TM = 6,000/10/10 ms) in different regions of the brain (cortex 3.9 \times 0.7 \times 3.2 mm³, corpus callosum 2.5 \times 1.0 \times 2.0 mm³, striatum 1.2 \times 1.4 \times 2.0 mm³, cerebellum 4 \times 3 \times 4 mm³) was performed at 9.4 T (Bruker Biospin GmbH). T2-weighted MRI (2D FSE, TR/TE = 4,200/43 ms, 8 echoes, resolution 100 \times 100 μ m², slice thickness 300 μ m) in axial and horizontal orientation served to ensure a proper position of volumes-of-interest. These images were further used to measure the ventricular volume by manual segmentation (Amira Software, Visage Imaging, Berlin, Germany). Metabolite quantification involved spectral evaluation by LCModel³³ and calibration with brain water concentration. Metabolites with Cramer-Rao lower bounds above 20% were excluded from further analyses unless noted otherwise.

29. Jaegle, M. *et al.* The POU proteins Brn-2 and Oct-6 share important functions in Schwann cell development. *Genes Dev.* **17**, 1380–1391 (2003).
30. Leone, D. P. *et al.* Tamoxifen-inducible glia-specific Cre mice for somatic mutagenesis in oligodendrocytes and Schwann cells. *Mol. Cell. Neurosci.* **22**, 430–440 (2003).
31. Minichiello, L. *et al.* Essential role for TrkB receptors in hippocampus-mediated learning. *Neuron* **24**, 401–414 (1999).
32. Michaelis, T., Boretius, S. & Frahm, J. Localized proton MRS of animal brain *in vivo*: Models of human disorders. *Prog. Nucl. Magn. Reson. Spectrosc.* **55**, 1–34 (2009).

Sustained translational repression by eIF2 α -P mediates prion neurodegeneration

Julie A. Moreno¹, Helois Radford¹, Diego Peretti¹, Joern R. Steinert¹, Nicholas Verity¹, Maria Guerra Martin¹, Mark Halliday¹, Jason Morgan¹, David Dinsdale¹, Catherine A. Ortori², David A. Barrett², Pavel Tsaytler³, Anne Bertolotti³, Anne E. Willis¹, Martin Bushell¹ & Giovanna R. Mallucci¹

The mechanisms leading to neuronal death in neurodegenerative disease are poorly understood. Many of these disorders, including Alzheimer's, Parkinson's and prion diseases, are associated with the accumulation of misfolded disease-specific proteins. The unfolded protein response is a protective cellular mechanism triggered by rising levels of misfolded proteins. One arm of this pathway results in the transient shutdown of protein translation, through phosphorylation of the α -subunit of eukaryotic translation initiation factor, eIF2. Activation of the unfolded protein response and/or increased eIF2 α -P levels are seen in patients with Alzheimer's, Parkinson's and prion diseases^{1–4}, but how this links to neurodegeneration is unknown. Here we show that accumulation of prion protein during prion replication causes persistent translational repression of global protein synthesis by eIF2 α -P, associated with synaptic failure and neuronal loss in prion-diseased mice. Further, we show that promoting translational recovery in hippocampi of prion-infected mice is neuroprotective. Overexpression of GADD34, a specific eIF2 α -P phosphatase, as well as reduction of levels of prion protein by lentivirally mediated RNA interference, reduced eIF2 α -P levels. As a result, both approaches restored vital translation rates during prion disease, rescuing synaptic deficits and neuronal loss, thereby significantly increasing survival. In contrast, salubrinal, an inhibitor of eIF2 α -P dephosphorylation⁵, increased eIF2 α -P levels, exacerbating neurotoxicity and significantly reducing survival in prion-diseased mice. Given the prevalence of protein misfolding and activation of the unfolded protein response in several neurodegenerative diseases, our results suggest that manipulation of common pathways such as translational control, rather than disease-specific approaches, may lead to new therapies preventing synaptic failure and neuronal loss across the spectrum of these disorders.

Neurodegenerative diseases pose an ever-increasing challenge for society and health care systems worldwide, but their molecular pathogenesis is still largely unknown and no curative treatments exist. Alzheimer's (AD), Parkinson's (PD) and prion diseases are separate clinical and pathological conditions, but it is likely they share common mechanisms leading to neuronal death. Mice with prion disease show misfolded prion protein (PrP) accumulation and develop extensive neurodegeneration (with profound neurological deficits), in contrast to mouse models of AD or PD, in which neuronal loss is rare. Uniquely therefore, prion-infected mice allow access to mechanisms linking protein misfolding with neuronal death. Prion replication involves the conversion of cellular PrP, PrP^C, to its misfolded, aggregating conformer, PrP^{Sc}, a process leading ultimately to neurodegeneration⁶. We have previously shown rescue of neuronal loss and reversal of early cognitive and morphological changes in prion-infected mice by depleting PrP in neurons, preventing prion replication and abrogating neurotoxicity^{7–9}. However, the molecular mechanisms underlying both the progression of disease, and those underlying recovery in PrP-depleted animals, were unknown.

To understand these processes better, we now analysed the evolution of neurodegeneration in prion-diseased mice. We examined hippocampi from prion-infected tg37 mice used in our previous experiments^{7–10}, in which the time course of impairment and recovery are clearly defined. Hemizygous tg37 mice express mouse PrP at approximately three times wild-type levels and succumb to Rocky Mountain Laboratory (RML) prion infection within 12 weeks post infection (w.p.i.)¹⁰. They first develop behavioural signs with decreased burrowing activity at approximately 9 w.p.i., after reduction in hippocampal synaptic transmission and first neuropathological changes^{7,8}. This is the window of reversibility when diseased neurons can still be rescued: PrP depletion up to 9 w.p.i., but not later, rescues neurotoxicity, as by 10 w.p.i. neuronal loss is established^{7–9}. We measured PrP levels, synapse number, levels of synaptic proteins and synaptic transmission in prion-infected mice weekly from 5 w.p.i., and burrowing behaviour from 6 w.p.i. We examined brains histologically and counted CA1 neurons. (Cohorts of at least 30 animals were used per group; biochemical and histological analyses were done on three mice per time point, burrowing behaviour on 12, *n* for other analyses is indicated in figure legends.) We found an early decline in synapse number in asymptomatic animals at 7 w.p.i. to approximately 55% of control levels (Fig. 1a), despite unchanged levels of several pre- and post-synaptic marker proteins (Fig. 1b). Reduced synapse number with normal synaptic protein levels is likely to reflect impaired structural plasticity of synapses at this early stage of disease. At 9 w.p.i., however, there was a sudden decline in synaptic protein levels to approximately 50% of control levels for several pre- (SNAP-25 and VAMP-2) and postsynaptic (PSD-95 and NMDAR1) proteins (Fig. 1b and Supplementary Fig. 1b). This was associated with further decline in synapse number, and the critical reduction in synaptic transmission, both in amplitude of evoked excitatory postsynaptic currents (EPSCs) and in the number of spontaneous miniature EPSCs (mEPSCs) in CA1 neurons (Fig. 1c and Supplementary Fig. 1e). This was coincident with behavioural change (Fig. 1d) and first spongiform pathology (Supplementary Fig. 1d), and was rapidly followed by the onset of neurodegeneration, resulting in 50% reduction in hippocampal pyramidal neurons at 10 w.p.i. (Fig. 1e). All animals developed overt motor signs and were terminally sick by 12 w.p.i.

The abrupt loss of synaptic proteins at 9 w.p.i. appeared to be a critical factor in the evolution of disease, occurring when synapse number and transmission were already declining. This could result from increased degradation, or decreased synthesis. Prion infection in mice is known to impair the ubiquitin proteasome system, causing reduction—not increase—in protein degradation¹¹. We therefore asked if protein synthesis was reduced through translational control mechanisms. Given that total PrP levels rise during disease (Fig. 2a), and that PrP is synthesized in the endoplasmic reticulum, we examined the translational repression pathway of the unfolded protein response (UPR). Rising levels of unfolded proteins detected

¹MRC Toxicology Unit, Hodgkin Building, University of Leicester, Lancaster Road, Leicester LE1 9HN, UK. ²Centre for Analytical Bioscience, School of Pharmacy, University of Nottingham, Nottingham NG7 2RD, UK. ³MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

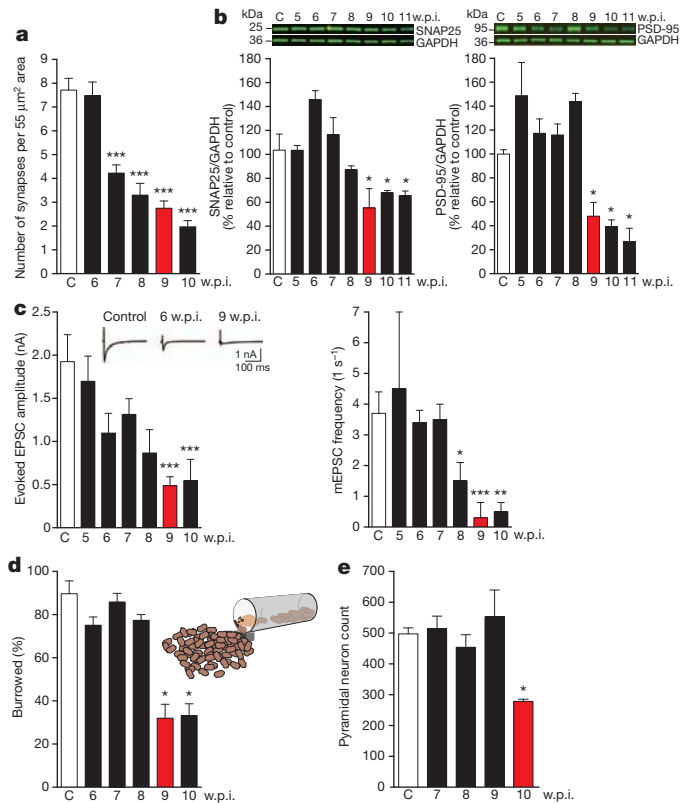


Figure 1 | Sudden decline of synaptic proteins is the key event leading to synaptic transmission failure and neuronal death in prion-diseased mice. **a**, Synapse number in the stratum radiatum of the hippocampal CA1 region (imaged by electron microscopy; Supplementary Fig. 1a), declined from 7 w.p.i. ($n = 2$ mice; 32 sections per time point; $***P < 0.0001$). **b**, Levels of pre- (SNAP-25) and postsynaptic (PSD-95) proteins measured relative to GAPDH declined abruptly at 9 w.p.i. Representative western blots are shown ($n = 3$ mice per time point, $*P = 0.05$, Student's t -test; two tails). **c**, Whole-cell recordings from CA1 neurons showed significant reduction at 9 w.p.i. in both amplitude of evoked EPSCs and frequency of spontaneous mEPSCs. Representative raw traces of evoked EPSCs are shown (left panel, inset). (For **c**, $n = 2$ mice, four to eight cells per time point. $*P < 0.05$; $**P < 0.005$; $***P < 0.0001$). **d**, Burrowing behaviour declines abruptly at 9 w.p.i. ($n = 12$ mice; $**P < 0.001$, Student's t -test; two tails). **e**, Number of CA1 pyramidal neurons is reduced by approximately 50% at 10 w.p.i. ($n = 3$ mice, three sections per mouse; $*P = 0.04$). All data show mean \pm s.e.m. One-way ANOVA with Tukey's post-test was used unless otherwise stated. Control mice were injected with normal brain homogenate and examined at each time point. Data from controls at all time points was averaged, owing to lack of variability over the time course, to simplify figures. For electrophysiological recordings and electron microscopy a single control time point at 10 w.p.i. was used.

by BiP/Grp78 (BiP) in the endoplasmic reticulum membrane cause auto-phosphorylation of protein kinase-like endoplasmic reticulum kinase (PERK). PERK-P phosphorylates eIF2 α , which blocks the initiation step of translation, reducing new protein synthesis¹². eIF2 α -P then induces ATF4 and CHOP expression, ultimately leading to caspase-12 cleavage, and expression of GADD34, the stress-induced eIF2 α -P-specific phosphatase and key effector of a negative feedback loop that terminates UPR signalling, allowing translational recovery (Fig. 2g).

Upregulation of various steps in the pathway are seen in human prion cases^{13,14} and in prion-infected mice¹³; and increased phosphorylation of eIF2 α occurs in AD and PD¹⁻⁴. We characterized this pathway in prion-diseased mice. We found that PERK-P and eIF2 α -P increased throughout the course of disease (Fig. 2b, c and Supplementary Fig. 2b–d), in parallel with rising levels of total PrP, and the presence of detectable protease-resistant PrP^{Sc} (Fig. 2a). GADD34 levels did not change, despite rising eIF2 α -P levels, suggesting insufficient GADD34

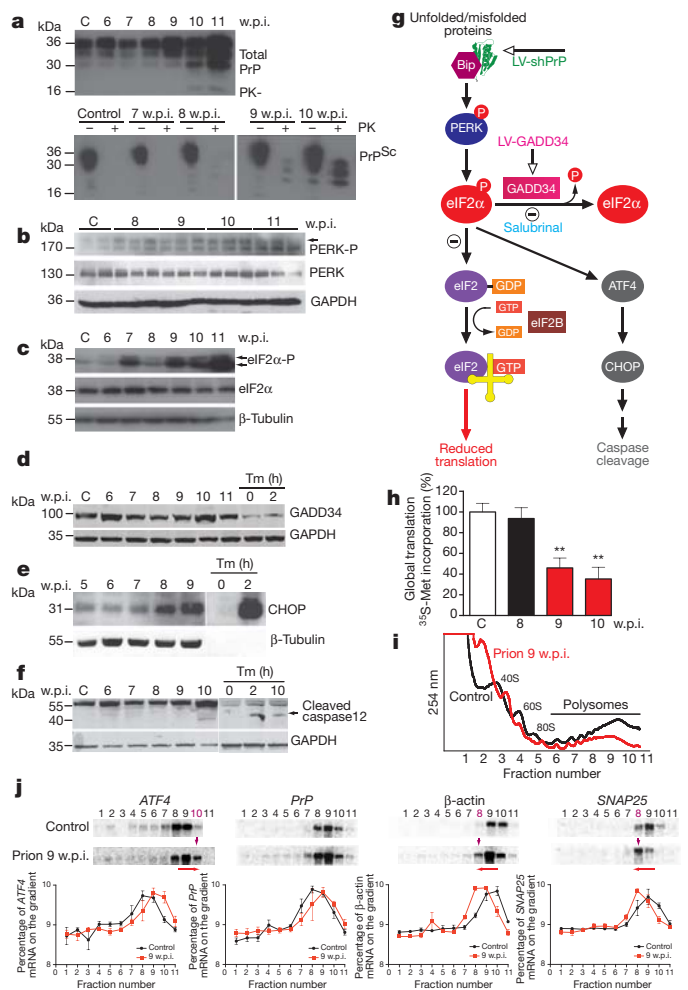


Figure 2 | Prion replication induces the UPR and results in eIF2 α -P-mediated translational repression. **a**, Total PrP and PrP^{Sc} levels (detected by addition of proteinase K (PK)) increase during prion infection; **b**, PERK-P and **c**, eIF2 α -P levels, rise during disease; **d**, GADD34 levels do not change but **e**, CHOP levels increase, throughout disease and **f**, caspase-12 is cleaved at 10 w.p.i. **g**, Scheme depicting the translational repression pathway of the UPR showing points for intervention by LV-shPrP, LV-GADD34 and salubrinal. Activation of the pathway results in reduction of global translation at 9 w.p.i. (**h**) determined by [³⁵S]methionine incorporation into hippocampal slices ($n = 3$ mice, six slices; $**P = 0.003$). **i**, Polysomal profiles from hippocampi show a reduction in active polysomes in fractions 6–11 from prion-infected mice at 9 w.p.i. **j**, Northern blots on polysomal fractions show increased translation of ATF4 with shift into fraction 10, and decreased translation of SNAP-25 and β -actin (shift into fraction 8) in prion-infected mice at 9 w.p.i. PrP translation is essentially unchanged. Quantitative line plots of northern blots show percentage of mRNA in each fraction found on the gradients. Tunicamycin (Tm)-treated HeLa cells were analysed at 0, 2 and 10 h as a control for UPR activation. All data show mean \pm s.e.m. One-way ANOVA with Tukey's post-test was used unless otherwise stated. Control mice at 11 w.p.i. are shown on western blots, and at 9 w.p.i. for northern blots and [³⁵S]methionine labelling; $n = 3$ mice for all experiments. For quantification of western blots, see Supplementary Fig. 2.

for dephosphorylation of increased eIF2 α -P (Fig. 2d and Supplementary Fig. 2e). Caspase-12 cleavage occurred at 10 w.p.i., following rising CHOP expression (Fig. 2e, f), coincident with onset of neuronal loss (Fig. 1e; see also Hetz *et al.*¹³). However, the exact effector mechanism of neuronal death is unclear: we found neither apoptosis, nor autophagy nor necrosis on examination of hippocampal slices (Supplementary Fig. 3); and neither Bax deletion, nor Bcl-2 overexpression¹⁵ nor caspase-12 deficiency¹⁶ is neuroprotective in prion disease.

We asked what effects the marked rise in eIF2 α -P levels at 9 w.p.i. had on overall protein synthesis in hippocampi. We found abrupt,

significant reduction in global translation rates, with a 50% decline in [35 S]methionine incorporation in hippocampal slices from prion-infected mice at 9 w.p.i. compared with mice at 8 w.p.i. and uninfected controls (Fig. 2h), confirming sudden onset of reduced protein synthesis. We also looked at translation of specific messenger RNA (mRNAs). We extracted polysomes from hippocampi of prion-infected mice. In successive polysomal fractions, mRNAs are associated with

increased numbers of actively translating ribosomes. The change from a single fraction to the next reflects a large change in translation rate for any specific mRNA. Polysomal profiles at 9 w.p.i. showed a reduction in the overall number of actively translating ribosomes, represented by the smaller area under the curve between fractions 6–11 in prion-infected mice (Fig. 2i). Northern blots for specific mRNAs in individual polysomal fractions confirmed changes in actively translated messages

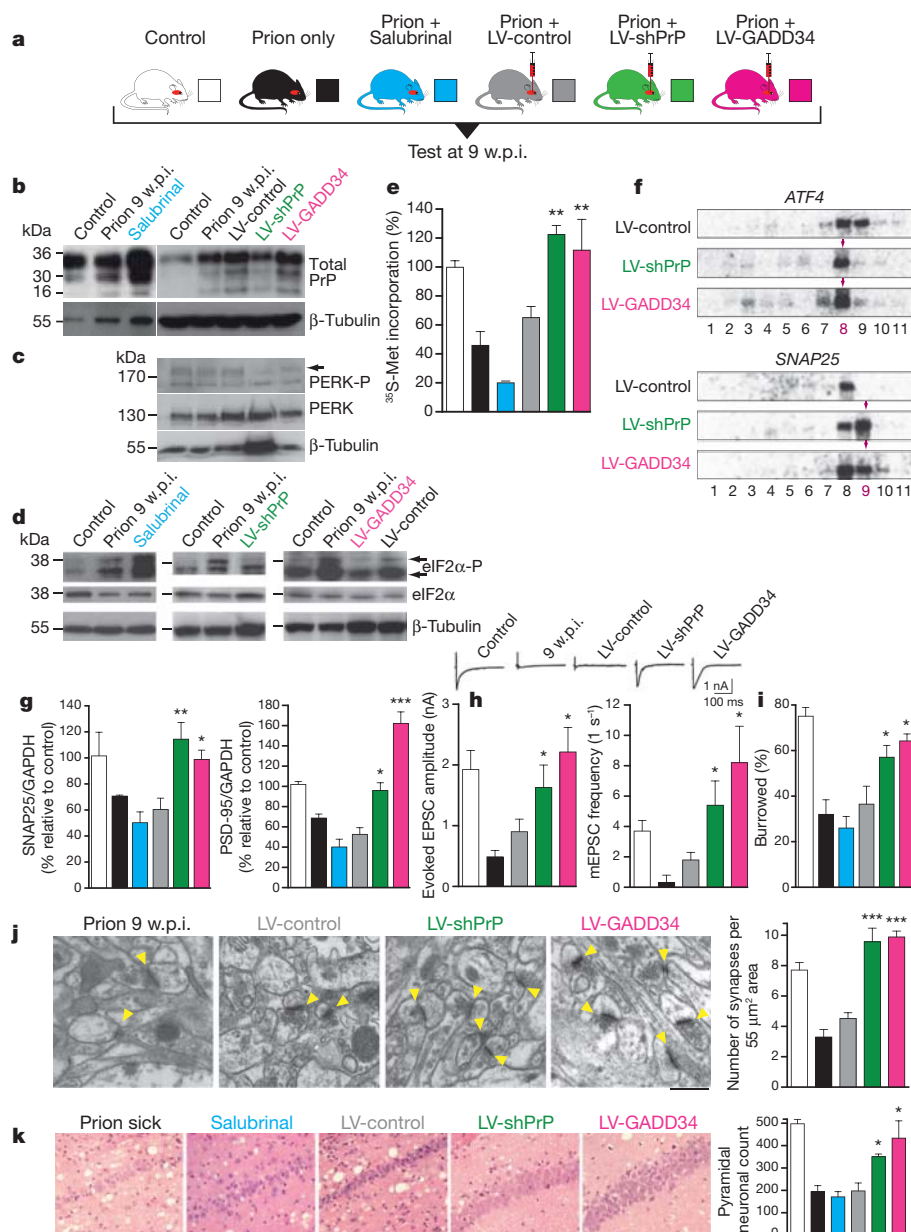


Figure 3 | Preventing eIF2 α -P formation or promoting its dephosphorylation in prion-diseased mice rescues synaptic failure and neuronal loss, while increased eIF2 α -P levels exacerbate neurotoxicity. **a**, Mice were infected with RML prions and treated with salubrinol (blue) or stereotactically injected with lentiviruses expressing anti-PrP shRNA (LV-shPrP; green), GADD34 (LV-GADD34; pink) or no insert (LV-control; grey) into both hippocampi. Control groups received no virus (prion only; black) or normal brain homogenate (control; white). Mice were tested at 9 w.p.i. **b**, LV-shPrP reduced total PrP and prevented UPR induction, reducing levels of PERK-P (**c**) and eIF2 α -P (**d**). LV-GADD34 reduced eIF2 α -P despite PERK-P induction, and salubrinol increased eIF2 α -P. **e**, Both LV-GADD34 and LV-shPrP prevented reduction in global translation at 9 w.p.i., but salubrinol reduced translation rates even further. **f**, LV-GADD34 and LV-shPrP reversed prion-induced eIF2 α -P-mediated translational changes of specific mRNAs in polysomal fractions shown on northern blots. **g**, Synaptic protein levels,

h, synaptic transmission, **i**, burrowing behaviour and **j**, synapse number were protected by GADD34 treatment and PrP knockdown. Salubrinol exacerbated protein loss. Representative electron micrographs (arrowheads denote individual synapses) and quantification are shown ($n = 2$ mice, 32 slices per mouse for each analysis). **k**, LV-GADD34 and LV-shPrP resulted in extensive neuroprotection of hippocampal CA1 pyramidal neurons and spongiosis (haematoxylin and eosin stained sections) and chart, right, when the animals were dying of scrapie at nearly 14 w.p.i. Salubrinol accelerated neurodegeneration with extensive neuronal loss seen at 9 w.p.i., earlier than in prion-sick animals at 12 w.p.i. Scale bar, 50 μ m and 2 μ m for electron micrographs. All data in bar charts show mean \pm s.e.m. One-way ANOVA with Tukey's post-test was used for multiple comparisons; * $P < 0.05$; ** $P < 0.005$; *** $P < 0.0001$. For all experiments $n = 3$ mice, unless otherwise stated. All controls were at 9 w.p.i. For quantification of Northern and western blots, see Supplementary Figs 4 and 9.

consistent with eIF2 α -P induction. Thus SNAP-25 and β -actin mRNAs showed a left shift to a lower polysomal fraction (Fig. 2j), representing reduced active translation (Fig. 1b). In contrast, *ATF4* mRNA (which escapes eIF2 α -P-mediated inhibition of translation, owing to the presence of upstream open reading frames in its 5' untranslated region (UTR)^{17,18}), showed increased active translation, represented by a right shift to a higher polysomal fraction (Fig. 2j and Supplementary Figs 4 and 5). *PrP* mRNA did not show reduced translation, possibly because of the existence of similar translational control elements within the *PrP* gene. Indeed, human *PrP* mRNA has multiple upstream AUG upstream open reading frames in its 5' UTR, which could allow it to escape eIF2 α -P translational inhibition in the same way as *ATF4* does^{17,18} (Supplementary Fig. 6).

Overall, these findings confirm that reduction in protein synthesis in prion disease is controlled at the translational, not the transcriptional, level, as it is rates of translation, not levels of total mRNA, that change (Supplementary Figs 1c and 2g, h).

We propose that the key trigger to prion neurodegeneration is the continued, unchecked activation of the UPR due to rising levels of PrP during disease, with fatal repression of translation rates. Importantly, prion neurotoxicity relates in a dose-dependent manner to PrP expression^{19–21}. We therefore asked if levels of eIF2 α -P and onset of neurodegeneration were related to levels of PrP in different strains of mice. We found that in homozygous tg37 mice, which overexpress PrP approximately sixfold, eIF2 α -P was induced at 6 w.p.i., and mice succumbed to prion infection at approximately 8 w.p.i. In wild-type C57/BL6 mice, which express 1 \times levels of PrP, eIF2 α -P was induced at 16 w.p.i. and animals succumbed at approximately 22 w.p.i. Thus, as for hemizygous tg37 mice where PrP was expressed at three times wild-type levels and eIF2 α -P was induced at 9 w.p.i. followed by death at 12 w.p.i. (Fig. 2c), in each case there was a corresponding critical decline in synaptic proteins and synapse number after eIF2 α -P induction (Supplementary Fig. 7 and Fig. 1a, b).

Transient eIF2 α phosphorylation is beneficial to cells overloaded with misfolded proteins: it reduces protein synthesis and increases availability of chaperones, promoting refolding^{22,23}. However, persistently high levels of eIF2 α -P are detrimental *in vitro*²⁴. To test directly the role of eIF2 α -P in prion neurodegeneration *in vivo*, we first asked if reduction of eIF2 α -P levels in prion disease would be neuroprotective. We used two approaches. We overexpressed GADD34, the eIF2 α -P-specific phosphatase to reduce eIF2 α -P levels directly. In a separate experiment, we used targeted RNA interference of PrP to abrogate UPR activation and prevent eIF2 α -P formation. We then asked if increased levels of eIF2 α -P exacerbate prion neurotoxicity by using salubrinal, a specific small-molecule inhibitor of eIF2 α -P dephosphorylation⁵, in infected mice. Salubrinal penetrates the blood–brain barrier (Supplementary Fig. 8) and has been used for modulation of eIF2 α -P-dependent effects in endoplasmic reticulum stress-mediated processes in the central nervous system *in vivo* after peripheral administration^{25–27}. Mice were inoculated with prions and received hippocampal injections of lentiviruses expressing GADD34 (LV-GADD34), anti-PrP shRNA (LV-shPrP) or yellow fluorescent protein (YFP) only (LV-control) at 5 w.p.i., allowing 4 weeks for lentiviral expression to occur, before testing the effects of treatment on eIF2 α -P levels and neurotoxicity at 9 w.p.i. (All virally expressed constructs were driven by the CAMKII promoter for neuron-specific expression (Supplementary Fig. 9a–c)). Another group of prion-infected mice received daily intraperitoneal injections of salubrinal (1 mg kg^{–1}), for 1 week, from 8 w.p.i., with controls receiving vehicle alone. Two further control groups received normal brain homogenate, or RML prion inoculation alone (Fig. 3a).

We examined mice from each group at 9 w.p.i. when eIF2 α -P-mediated translational repression occurs (Fig. 2c, h). LV-GADD34 treatment did not reduce PrP levels (Fig. 3b) and PERK-P levels were equivalent to those in prion-only or LV-control treated animals (Fig. 3c and Supplementary Fig. 9d), confirming UPR activation in

these mice. Critically, however, eIF2 α -P levels were reduced (Fig. 3d and Supplementary Fig. 9e), strongly supporting its dephosphorylation by lentivirally mediated GADD34 expression. LV-shPrP treatment reduced PrP levels (Fig. 3b) and prevented the PrP-induced rise in PERK-P and eIF2 α -P seen in untreated animals (Fig. 3c, d), confirming prevention of UPR activation. Both GADD34 overexpression and PrP knockdown prevented prion-induced eIF2 α -P-mediated translational repression, with restoration of global rates at 9 w.p.i. (Fig. 3e) and prevention of eIF2 α -P-induced changes in translation of specific mRNAs (Fig. 3f and Supplementary Fig. 4). As a result, synaptic protein levels, synaptic transmission and synapse number in prion-diseased mice treated with GADD34 or PrP knockdown were protected and equivalent to levels in uninfected control mice (Fig. 3g, h, j). Burrowing deficits were prevented (Fig. 3i) and there was extensive neuronal protection in the hippocampus, with no neuronal loss and markedly reduced spongiform change (Fig. 3k). Further, targeted expression of LV-GADD34 and focal PrP knockdown had a modest, but highly significant, effect on survival, increasing this to 90 \pm 3 days and 92 \pm 5 days respectively, compared with 83 \pm 2 days for prion-only mice (Fig. 4), and to 82 \pm 2 days for LV-control injected mice (Supplementary Fig. 10). For both LV-GADD34 and LV-shPrP, treatment was localized to the dorsal hippocampus, a very small area of the brain, so prion infection in the rest of the brain was fatal, but neuroprotection in GADD34-treated animals was seen even when the animals were terminally sick (Fig. 3k). More extensive brain-wide delivery of GADD34, or targeting of this pathway, would be predicted to increase survival further and give more widespread neuroprotection. Critically, treatment with salubrinal had the opposite effect, by preventing dephosphorylation of eIF2 α -P. Thus, eIF2 α -P levels were markedly higher at 9 w.p.i. than in prion-only controls (Fig. 3d and Supplementary Fig. 9e), causing further repression of global translation (Fig. 3e) and reduction of synaptic proteins (Fig. 3g). Salubrinal treatment resulted in earlier severe neuronal loss (Fig. 3k), and significantly accelerated disease, compared with untreated prion-infected mice (Fig. 4).

In conclusion, we have shown that PrP replication causes sustained UPR induction with persistent, deleterious expression of eIF2 α -P in prion disease. The resulting chronic blockade of protein synthesis leads to synaptic failure, spongiosis and neuronal loss. Promoting eIF2 α -P dephosphorylation rescues vital translation rates and is thereby neuroprotective, whereas preventing this further reduces translation and enhances neurotoxicity. The data support the development of generic proteostatic approaches^{22,28} to therapy—fine-tuning protein synthesis—in prion, and perhaps other neurodegenerative disorders involving protein misfolding.

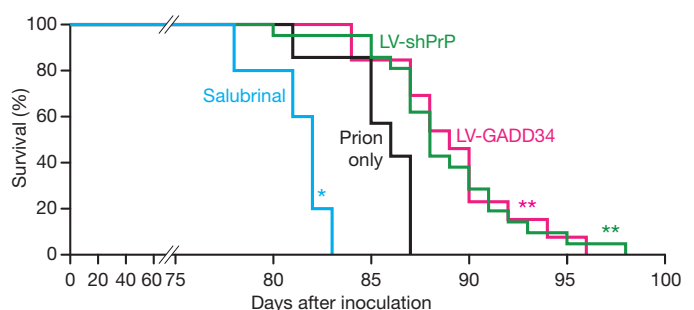


Figure 4 | Reducing eIF2 α -P levels in prion-diseased mice significantly increases survival. Kaplan–Meier survival plots for prion-infected mice treated with salubrinal ($n = 10$), prions alone ($n = 7$), hippocampal injections of LV-GADD34 ($n = 13$) or LV-shPrP ($n = 14$). Focal treatment with LV-GADD34 and LV-shPrP resulted in significantly increased survival compared with prion-infected mice with no treatment (** $P = 0.007$; Student's t -test); salubrinal treatment reduced survival (* $P = 0.03$). For plots for LV-control and DMSO controls, see Supplementary Fig. 10.

METHODS SUMMARY

Prion infection of mice. All animal work conformed to UK regulations and institutional guidelines, performed under Home Office guidelines. tg37 (ref. 10) and C57/Bl6N mice (Harlan) were inoculated with 1% brain homogenate of Chandler/RML prions aged 3–4 weeks, as described⁷. Animals were culled when they developed clinical signs of scrapie. Control mice received 1% normal brain homogenate. Hippocampi were processed for protein, RNA or histological analysis^{7–10}. For all analyses $n = 3$ mice unless otherwise stated.

Lentiviruses and stereotaxic surgery. Lentiviral plasmids were generated using the Invitrogen Gateway cloning system²⁹. The neuron-specific promoter CAMKII was used to drive shPrP expression, carboxy (C)-terminal GADD34 expression, or YFP alone (control virus). Viruses were injected stereotactically into the CA1 region of the hippocampus as described⁹.

Salubrial treatment. Mice received daily intraperitoneal injections of 1 mg kg⁻¹ of salubrial (Calbiochem), or vehicle (diluted dimethylsulphoxide (DMSO) in saline (Sigma))²⁵, for 7 days from 8 w.p.i.

Experimental analyses. Synapse numbers were counted in electron micrographs of the stratum radiatum of the hippocampal CA1 region³⁰. Synaptic marker proteins, UPR pathway proteins and PrP were analysed by immunoblotting of brain homogenates. PrP^{Sc} was detected after proteinase K digestion⁷. Whole-cell recordings were done in acute hippocampal slices to measure synaptic transmission³¹. Global translation levels were detected using [³⁵S]methionine incorporation in acute hippocampal slices; translations of specific transcripts in polysomal fractions from hippocampi were analysed by northern blotting³². Neuronal counts were determined by quantifying NeuN-positive pyramidal CA1 neurons⁹. All analyses were performed using hippocampi from three mice in triplicate unless otherwise stated. Burrowing behaviour was performed as described on groups of 10 or more mice⁸. Statistical analyses were performed using GraphPad/Prism software, version 5, using Student's *t*-test for data sets with normal distribution and a single intervention. Analysis of variance (ANOVA) testing was performed using one-way analysis with Tukey's post-hoc test for multiple comparisons.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 October 2011; accepted 16 March 2012.

Published online 6 May 2012.

- Hoozemans, J. J. *et al.* Activation of the unfolded protein response in Parkinson's disease. *Biochem. Biophys. Res. Commun.* **354**, 707–711 (2007).
- Hoozemans, J. J. *et al.* The unfolded protein response is activated in pretangle neurons in Alzheimer's disease hippocampus. *Am. J. Pathol.* **174**, 1241–1251 (2009).
- Hoozemans, J. J. *et al.* The unfolded protein response is activated in Alzheimer's disease. *Acta Neuropathol.* **110**, 165–172 (2005).
- Unterberger, U. *et al.* Endoplasmic reticulum stress features are prominent in Alzheimer disease but not in prion diseases *in vivo*. *J. Neuropathol. Exp. Neurol.* **65**, 348–357 (2006).
- Boyce, M. *et al.* A selective inhibitor of eIF2 α dephosphorylation protects cells from ER stress. *Science* **307**, 935–939 (2005).
- Prusiner, S. B. Molecular biology of prion diseases. *Science* **252**, 1515–1522 (1991).
- Mallucci, G. *et al.* Depleting neuronal PrP in prion infection prevents disease and reverses spongiosis. *Science* **302**, 871–874 (2003).
- Mallucci, G. R. *et al.* Targeting cellular prion protein reverses early cognitive deficits and neurophysiological dysfunction in prion-infected mice. *Neuron* **53**, 325–335 (2007).
- White, M. D. *et al.* Single treatment with RNAi against prion protein rescues early neuronal dysfunction and prolongs survival in mice with prion disease. *Proc. Natl Acad. Sci. USA* **105**, 10238–10243 (2008).
- Mallucci, G. R. *et al.* Post-natal knockout of prion protein alters hippocampal CA1 properties, but does not result in neurodegeneration. *EMBO J.* **21**, 202–210 (2002).
- Kristiansen, M. *et al.* Disease-associated prion protein oligomers inhibit the 26S proteasome. *Mol. Cell* **26**, 175–188 (2007).
- Ron, D. & Walter, P. Signal integration in the endoplasmic reticulum unfolded protein response. *Nature Rev. Mol. Cell Biol.* **8**, 519–529 (2007).
- Hetz, C., Russelakis-Carneiro, M., Maundrell, K., Castilla, J. & Soto, C. Caspase-12 and endoplasmic reticulum stress mediate neurotoxicity of pathological prion protein. *EMBO J.* **22**, 5435–5445 (2003).
- Yoo, B. C. *et al.* Overexpressed protein disulfide isomerase in brains of patients with sporadic Creutzfeldt-Jakob disease. *Neurosci. Lett.* **334**, 196–200 (2002).
- Steele, A. D. *et al.* Diminishing apoptosis by deletion of Bax or overexpression of Bcl-2 does not protect against infectious prion toxicity *in vivo*. *J. Neurosci.* **27**, 13022–13027 (2007).
- Steele, A. D. *et al.* Prion pathogenesis is independent of caspase-12. *Prion* **1**, 243–247 (2007).
- Harding, H. P. *et al.* Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell* **6**, 1099–1108 (2000).
- Spriggs, K. A., Bushell, M. & Willis, A. E. Translational regulation of gene expression during conditions of cell stress. *Mol. Cell* **40**, 228–237 (2010).
- Bueler, H. *et al.* Mice devoid of PrP are resistant to scrapie. *Cell* **73**, 1339–1347 (1993).
- Manson, J. C., Clarke, A. R., McBride, P. A., McConnell, I. & Hope, J. PrP gene dosage determines the timing but not the final intensity or distribution of lesions in scrapie pathology. *Neurodegeneration* **3**, 331–340 (1994).
- Sandberg, M. K., Al-Doujaily, H., Sharps, B., Clarke, A. R. & Collinge, J. Prion propagation and toxicity *in vivo* occur in two distinct mechanistic phases. *Nature* **470**, 540–542 (2011).
- Tsaytler, P., Harding, H. P., Ron, D. & Bertolotti, A. Selective inhibition of a regulatory subunit of protein phosphatase 1 restores proteostasis. *Science* **332**, 91–94 (2011).
- Chafekar, S. M., Hoozemans, J. J., Zwart, R., Baas, F. & Scheper, W. A β _{1–42} induces mild endoplasmic reticulum stress in an aggregation state-dependent manner. *Antioxid. Redox Signal.* **9**, 2245–2254 (2007).
- Wiseman, R. L. *et al.* Flavonol activation defines an unanticipated ligand-binding site in the kinase-RNase domain of IRE1. *Mol. Cell* **38**, 291–304 (2010).
- Sokka, A. L. *et al.* Endoplasmic reticulum stress inhibition protects against excitotoxic neuronal injury in the rat brain. *J. Neurosci.* **27**, 901–908 (2007).
- Saxena, S., Cabuy, E. & Caroni, P. A role for motoneuron subtype-selective ER stress in disease manifestations of FALS mice. *Nature Neurosci.* **12**, 627–636 (2009).
- Zhu, Y. *et al.* Eif-2 α protects brainstem motoneurons in a murine model of sleep apnea. *J. Neurosci.* **28**, 2168–2178 (2008).
- Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916–919 (2008).
- White, M. D., Milne, R. V. & Nolan, M. F. A molecular toolbox for rapid generation of viral vectors to up- or down-regulate neuronal gene expression *in vivo*. *Front. Mol. Neurosci.* **4**, 8 (2011).
- Haustein, M. D. *et al.* Acute hyperbilirubinaemia induces presynaptic neurodegeneration at a central glutamatergic synapse. *J. Physiol. (Lond.)* **588**, 4683–4693 (2011).
- Steinert, J. R. *et al.* Nitric oxide is an activity-dependent regulator of target neuron intrinsic excitability. *Neuron* **71**, 291–305 (2011).
- Johannes, G. & Sarnow, P. Cap-independent polysomal association of natural mRNAs encoding c-myc, BiP, and eIF4G conferred by internal ribosome entry sites. *RNA* **4**, 1500–1513 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Read for imaging analysis, J. Edwards, T. Smith, J. McWilliam, P. Glynn, C. Molloy and University of Leicester, Department of Biological Services staff for technical assistance, J. Collinge (MRC Prion Unit) for the original RML prion inoculum, and K. Liddle for reading the manuscript. This work was funded by the Medical Research Council, UK.

Author Contributions J.A.M. did most of the experimental work and analysis. N.V. and M.G.M. performed stereotaxic surgery and prion inoculations. H.R., D.P., M.H. and J.M. performed various experiments, D.D. performed electron microscopy, J.R.S. performed electrophysiological analysis, C.A.O. and D.A.B. performed mass spectrometry analysis, P.T. and A.B. worked with J.A.M. in Cambridge, A.E.W. and M.B. contributed expertise and direction on translational control mechanisms, and G.R.M. directed and supervised the project. J.A.M. and G.R.M. wrote the paper. All authors contributed to discussion, analysis of data and the final draft of paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.R.M. (grm7@le.ac.uk).

METHODS

Prion infection of mice. All animal work conformed to UK regulations and institutional guidelines, performed under Home Office guidelines. tg37 (ref. 10) and C57/BL6N mice (Harlan) were inoculated with 1% brain homogenate of Chandler/RML prions aged 3–4 weeks, as described⁷. Animals were culled when they developed clinical signs of scrapie. Control mice received 1% normal brain homogenate. Hippocampi were processed for protein, RNA or histological analysis^{7–10}. For all analyses $n = 3$ mice unless otherwise stated.

Lentiviruses and stereotaxic surgery. Lentiviral plasmids were generated using the Invitrogen Gateway cloning system²⁹. The neuron-specific promoter CAMKII was used to drive shPrP expression, C-terminal GADD34 expression, or YFP alone (control virus). Viruses were injected stereotactically into the CA1 region of the hippocampus as described⁹.

Salubrin treatment. Mice received daily intraperitoneal injections of 1 mg kg⁻¹ of salubrin (Calbiochem), or vehicle (diluted DMSO in saline (Sigma))²⁵, for 7 days from 8 w.p.i.

Experimental analyses. Synapse numbers were counted in electron micrographs of the stratum radiatum of the hippocampal CA1 region³⁰. Synaptic marker proteins, UPR pathway proteins and PrP were analysed by immunoblotting of brain homogenates. PrP^{Sc} was detected after proteinase K digestion⁷. Whole-cell recordings were done in acute hippocampal slices to measure synaptic transmission³¹. Global translation levels were detected using [³⁵S]methionine incorporation in acute hippocampal slices; translations of specific transcripts in polysomal fractions from hippocampi were analysed by northern blotting³². Neuronal counts were determined by quantifying NeuN-positive pyramidal CA1 neurons⁹. All analyses were performed using hippocampi from three mice in triplicate unless otherwise stated. Burrowing behaviour was performed as described on groups of 10 or more mice⁸. Statistical analyses were performed using GraphPad/Prism software, version 5, using Student's *t*-test for data sets with normal distribution and a single intervention. ANOVA testing was performed using one-way analysis with Tukey's post-hoc test for multiple comparisons.

Generation of lentiviral plasmids. Lentiviral plasmids were generated by using the Invitrogen Gateway cloning system as described²⁹. ATT-flanked cassettes were constructed by PCR amplification using ATT recombination site sequences and 20–25 template-specific sequences. The C terminus GADD34 cassette was amplified from Flag-tagged C-term GADD34 using attB5 5'-GGGGACAACCTTGTATACAAAAGTTGGCACCATTGCGTTCAGGAGAGGCGTCCGA-3'; and attB4 5'-GGGGACAACCTTGTATAGAAAAGTTGGGTGTTGGTCTCAGCCACGCTCCAC-3' primers; WPRE and YFP cassettes were amplified from pLL3.7-shPrP plasmid using attB3 5'-GGGGACAACCTTGTATATAATAAAGTTGTCAACCTCTGGATTACAAAATTTGT-3' and attB2 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTATGCGGGGAGGCGGCCCAAGGGAGA-3' primers for WPRE and attB4r 5'-ACCATGGTGTGAGCAAGGGCGA-3' and attB3r 5'-TTACTTGTACAGCTCGTCCATGCCG-3' primers for YFP. The CAMKII promoter was amplified from mouse cDNA using attB1 5'-GGGGACAA GTTTGTACAAAAAGCAGGCTACTTGTGGACTAAGTTTGTTCGC-3' and attB5r primer 5'-GGGGACAACCTTTGTATACAAAGTTGTCTGCCCCAGA ACTAGGG-3' primers. Cassettes were amplified and recombined into appropriate pDONR entry vectors. Once the entry plasmids were generated they were recombined with the pLenti6/BLOCK-iT/DEST vector (Invitrogen) to construct lentiviral plasmids. We used lentiviral plasmids containing shPrP and empty vector constructs as described⁹. Lentiviruses were generated using DNA/Ca²⁺ phosphate transfection of HEK293 cells³³. Additional stocks of virus were generated by GenTarget and titre determined using fluorescence-activated cell sorting (BD FACS Calibur). Viruses were used with a final titre of 0.6×10^8 to 1.5×10^8 TU.

Stereotaxic injection. Mice were anaesthetized using isoflurane. Injections were into the CA1 region of the hippocampus as described⁹.

Electron microscopy. Semi-thin sections for electron microscopy were obtained by terminal perfusion of mice as described³⁰. Ultrathin sections (70 nm) were examined in a JEOL 100-CXII electron microscope (JEOL (UK)) equipped with a 'Megaview III' digital camera (Olympus Soft Imaging Solutions). Synapses were scored using criteria of structures showing a postsynaptic density, containing synaptic vesicles and a synaptic junction. Thirty-two images from two mice were used for scoring.

Electrophysiology. Whole-cell recordings were made from identified CA1 neurons and recording performed as described³⁰. In brief, neurons were voltage clamped

using a Multiclamp 700B amplifier and pClamp 10.3 software (Molecular Devices) and EPSCs were evoked by stimulation with bipolar platinum electrode at 37 °C. Pipettes (2.5–3.5 M Ω) were filled with a solution containing (in mM): KCl 110, HEPES 40, EGTA 0.2, MgCl₂ 1, CaCl₂ 0.1; pH was adjusted to 7.2 with KOH. Neurons were visualized with $\times 60$ objective lenses on a Nikon FS600 microscope fitted with differential interference contrast optics. Four to eight cells were measured per mouse in at least two animals per experiment.

Immunoblotting. Protein samples were isolated from hippocampi using protein lysis buffer (50 mM Tris, 150 mM NaCl, 2 mM EDTA, 1 mM MgCl₂, 100 mM NaF, 10% glycerol, 1% Triton X-100, 1% Na deoxycholate, 0.1% SDS and 125 mM sucrose) supplemented with Phos-STOP and protease inhibitors (Roche). Synaptic protein levels were determined by resolving 20 mg of protein on SDS-polyacrylamide gel electrophoresis gels, transferred onto nitrocellulose membrane and incubated with primary antibodies, SNAP-25, (1:10,000; Abcam), VAMP2 (1:5,000; Synaptic Systems), NMDA-R1 (1:1,000; Sigma) and PSD95 (1:1,000; Millipore). Odyssey IRDye800 secondary antibodies (1:5,000; LI-COR) were applied, visualized and quantified using Odyssey infrared imager (LI-COR; software version 3.0). Protein for PrP levels and UPR pathway activation were determined using the primary antibodies, 8H4 for total PrP (1:1,000; Abcam), ICSM35 for PrP^{Sc} (1:10,000; D-GEN), PERK-P, total PERK, P-eIF2 α , eIF2 α (1:1,000; Cell Signaling), CHOP (1:1,000; ThermoScientific), caspase-12 (1:1,000; Exalpha), BiP/Grp78 (1:1,000; Stressgen) and GADD34 (1:1,000; ProteinTech). Horseradish-peroxidase-conjugated secondary antibodies (1:5,000; DAKO) were applied and protein visualized using enhanced chemiluminescence (GE Healthcare) and quantified using ImageJ. Antibodies against GAPDH (1:5,000; Santa Cruz) or β -tubulin (1:5,000; Millipore) were used to determine loading.

Hippocampal slice preparation and [³⁵S]methionine labelling. Slices were dissected in an oxygenated cold (2–5 °C) sucrose artificial cerebrospinal fluid containing (in mM): 26 NaHCO₃, 2.5 KCl, 4 MgCl₂, 0.1 CaCl₂, and 250 sucrose. Hippocampal slices were prepared using a tissue chopper (McIlwain). Slices were allowed to recover in normal artificial cerebrospinal fluid buffer while being oxygenated at 37 °C for 1 h, incubated with [³⁵S] methionine label for 1 h, then homogenized. Proteins were precipitated with trichloroacetic acid; incorporation of radiolabel was measured by scintillation counting (Winspectral, Wallac).

Polysomal fraction preparation and northern blots. Sucrose density gradient centrifugation was used to separate hippocampal homogenates into polysomal and subpolysomal fractions. Polysomal fractions were isolated as described³². Briefly, hippocampi were dissected in ice-cold gradient buffer (0.3 M NaCl, 15 mM MgCl₂, 15 mM Tris-HCl (pH 7.4), 0.1 mg ml⁻¹ cyclohexamide and 1 mg ml⁻¹ heparin). The hippocampal tissue was homogenized in gradient buffer containing RNase inhibitors and 1.2% TritonX-100 added. Samples were centrifuged and the supernatants layered onto 10–60% sucrose gradients. The gradients were sedimented at 38,000 r.p.m. using a TH-641 rotor (Thermo Scientific) for 120 min at 4 °C. One millilitre fractions were collected from the gradients into 3 ml of 7.7 M guanidine-HCl using a Foxy R1 gradient fractionator (Teledyne ISCO; ISCO peak Trak version 1.10 software) with continuous measurement of the absorbance at 254 nm. RNA was then isolated and equal volumes of each fraction were analysed by northern blot analysis³².

Neuropathological examination of brain sections. Paraffin-embedded brains were either stained with haematoxylin and eosin (H&E) or incubated with NeuN antibody (1:200; Millipore) for neuronal counts. CA1 pyramidal neuron counts were determined using three serial sections from three separate mice. All images were taken using a Zeiss Axiovert 200M microscope with a Colibri illumination system with Axiovision 4.8 software (Zeiss) and NeuN positive cells were counted using the Volocity v.5 (Perkin Elmer) software.

Burrowing. Burrowing was performed as described. At least 10 mice were burrowed for each individual experimental group^{10,34}.

Statistical analysis. Student's *t*-tests were applied to all data sets with two tails (two samples; unequal variance). ANOVA testing was performed using one-way analysis with Tukey's post-hoc test for group effects. Statistical tests were performed using GraphPad/Prism version 5. All data in bar charts show mean \pm s.e.m.

33. Jiang, M., Deng, L. & Chen, G. High Ca²⁺-phosphate transfection efficiency enables single neuron gene analysis. *Gene Ther.* **11**, 1303–1311 (2004).

34. Cunningham, C. *et al.* Synaptic changes characterize early behavioural signs in the ME7 model of murine prion disease. *Eur. J. Neurosci.* **17**, 2147–2155 (2003).

Discovery of a sensory organ that coordinates lunge feeding in rorqual whales

Nicholas D. Pyenson^{1,2}, Jeremy A. Goldbogen³, A. Wayne Vogl⁴, Gabor Szathmary⁵, Richard L. Drake⁶ & Robert E. Shadwick⁷

Top ocean predators have evolved multiple solutions to the challenges of feeding in the water^{1–3}. At the largest scale, rorqual whales (*Balaenopteridae*) engulf and filter prey-laden water by lunge feeding⁴, a strategy that is unique among vertebrates¹. Lunge feeding is facilitated by several morphological specializations, including bilaterally separate jaws that loosely articulate with the skull^{5,6}, hyper-expandable throat pleats, or ventral groove blubber⁷, and a rigid y-shaped fibrocartilage structure branching from the chin into the ventral groove blubber⁸. The linkages and functional coordination among these features, however, remain poorly understood. Here we report the discovery of a sensory organ embedded within the fibrous symphysis between the unfused jaws that is present in several rorqual species, at both fetal and adult stages. Vascular and nervous tissue derived from the ancestral, anterior-most tooth socket insert into this organ, which contains connective tissue and papillae suspended in a gel-like matrix. These papillae show the hallmarks of a mechanoreceptor, containing nerves and encapsulated nerve termini. Histological, anatomical and kinematic evidence indicate that this sensory organ responds to both the dynamic rotation of the jaws during mouth opening and closure, and ventral groove blubber⁷ expansion through direct mechanical linkage with the y-shaped fibrocartilage structure. Along with vibrissae on the chin⁹, providing tactile prey sensation, this organ provides the necessary input to the brain for coordinating the initiation, modulation and end stages of engulfment, a paradigm that is consistent with unsteady hydrodynamic models and tag data from lunge-feeding rorquals^{10–13}. Despite the antiquity of unfused jaws in baleen whales since the late Oligocene¹⁴ (~23–28 million years ago), this organ represents an evolutionary novelty for rorquals, based on its absence in all other lineages of extant baleen whales. This innovation has a fundamental role in one of the most extreme feeding methods in aquatic vertebrates, which facilitated the evolution of the largest vertebrates ever.

Large marine suspension feeders have evolved many times over the past 200 million years (Myr)^{1–3}. These vertebrates, which include extinct bony fish, chondrichthyan and baleen whales (Mysticeti), all show similar morphological specializations for feeding in water while maintaining large body sizes^{2,3}. For those vertebrates with solely aquatic ancestries, such as bony fish and chondrichthyan, feeding modes mostly consist of variations on unidirectional ram filter feeding, using gill rakers¹. By contrast, baleen whales are secondarily adapted to marine environments from a terrestrial ancestry³. Their feeding specializations thus reflect a trade-off between evolutionary innovations (for example, baleen) and fundamental constraints from a mammalian body plan (such as air breathing)³. The origin of suspension feeding in mysticetes represents an evolutionary novelty with no parallel in any other mammalian or known synapsid lineage¹⁵. Moreover, the innovation of baleen preceded the origin of extremely large body sizes in mysticetes¹⁶, which rank among the largest vertebrates ever.

Although mysticetes use several different modes of filter feeding³, rorquals exclusively engulf large volumes of prey-laden water in a single, rapid gulp⁴. In fin whales (*Balaenoptera physalus*), for example, the entire process of a lunge occurs in a short time span (~6 s), during which approximately 60–80 m³ of water and prey are engulfed, a volume equal to or greater than that of the individual rorqual itself¹⁷ (Fig. 1). During a lunge, rorquals accelerate to high speed and open their jaws in a complex sequence of rotation around three orthogonal axes⁶, including an overall gape angle that approaches 90 degrees at the peak of the lunge⁵. Dynamic pressure imposed on the floor of the mouth forces inversion of the tongue¹⁸ and expansion of the ventral groove blubber (VGB) to accommodate the engulfed water^{5,7}. Hydro-mechanical models suggest that the expansion of the oropharyngeal cavity (or ventral pouch⁸) requires active resistance, in a highly coordinated fashion, by eccentric action from musculature associated with the VGB^{10–13}. The lunge sequence ends once the mandibles close around the volume of engulfed water, with the VGB and the oropharyngeal cavity slowly returning to its original size, allowing the baleen plates to filter the captured prey from the engulfed water³.

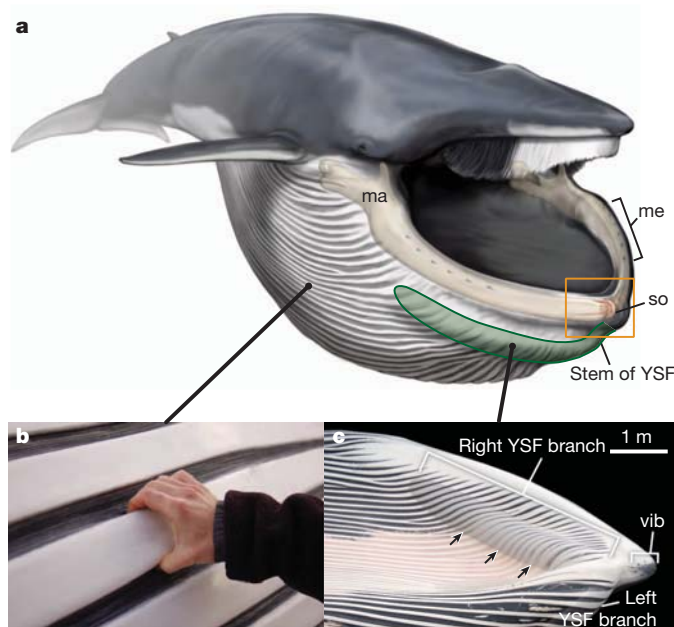


Figure 1 | Key jaw anatomy of a rorqual lunge, shown in a fin whale (*B. physalus*). **a**, Jaw in oblique view. **b**, Laterally extensible, accordion-like VGB. **c**, YSF embedded in the VGB of an adult fin whale, demarcated by arrows and digitally enhanced with a low opacity mask. ma, mandibles; me, mental foramina; so, sensory organ; vib, vibrissae. Art by C. Buell.

¹Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, PO Box 37012, Washington, District of Columbia 20013-7013, USA. ²Departments of Mammalogy and Paleontology, Burke Museum of Natural History and Culture, Seattle, Washington 98195, USA. ³Cascadia Research Collective, 218½ West 4th Avenue, Olympia, Washington 98501, USA. ⁴Department of Cellular and Physiological Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. ⁵FPIInnovations, 2665 East Mall, Vancouver, British Columbia V6T 1W5, Canada. ⁶Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, Ohio 44195, USA. ⁷Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

In addition to the accordion-like VGB, lunge feeding is facilitated by a suite of bony and soft tissue morphological adaptations^{5–9} (Fig. 1). The lower jaw consists of two mandibles unfused at the symphysis and flexibly anchored to the skull by fibrous joints^{5,6}. These atypical fibrous temporomandibular joints⁵ increase kinetic freedom and resist shear forces during the rapid excursions of a lunge. The mandibular symphysis has a direct mechanical linkage to the VGB⁷ through the y-shaped fibrocartilage (YSF) structure⁸, which has a raised stem that emerges from the symphysis and extends posteriorly in two branches, parallel to the jaws and embedded within the VGB⁸. How these structures are integrated into the engulfment apparatus has remained enigmatic, although hydro-mechanical models and kinematic data from tagged rorquals suggest that lunge feeding is a highly coordinated process^{10–13}. Here, we identify a novel sensory organ that is mechanically linked to both bony and soft tissues, and can provide the brain with mechanosensory information for coordinating the rapid and marked expansion of the oral cavity during a lunge.

We analysed the functional morphology of the mandibular symphysis in small, medium and large balaenopterid species, including both adult and fetal specimens (see Supplementary Information). Midline sagittal cuts along the mandibular symphysis showed that the sensory organ is mostly located on the dorsal half of the region between the symphyseal surfaces of the mandibles, and occupies a roughly spheroidal cavity. The organ is bound on all sides by dense connective tissue; the inner surface of the cavity is lined by a discontinuous squamous layer of cells and filled with a viscous gel-like matrix. The gel-like matrix supports numerous small connective tissue papillae that extend from the inner surface and sometimes cluster in irregular numbers within the cavity. Notably, the papillae contain nerves and encapsulated nerve termini (Fig. 2).

We also analysed the three-dimensional soft-tissue anatomy of an adult fin whale mandibular symphysis before dissection, using both magnetic resonance imaging (MRI) and X-ray computed tomography (XRCT) scanning (Fig. 3). Our analyses showed that the organ receives neurovascular bundles that emerge from vestigial alveolar foramina of

the mandibles, a structure that is homologous with the lower first incisor tooth socket of fossil ‘toothed’ mysticetes^{14–16}. The presence and identity of these neurovascular bundles, which contain branches of the mandibular division of the trigeminal nerve, were confirmed by subsequent dissection of the same specimen. They were also present in fetal specimens of the same species (see Supplementary Information). Three-dimensional renderings also indicated that these bundles are asymmetrically arranged, with a single bundle extending medially from one mandible. We verified that this asymmetry fluctuates in dominance, with 60% from the left in fin whales ($n = 10$). In minke whales (*Balaenoptera acutorostrata*), we observed bundle dominance from both the left and right ($n = 2$). XRCT and MRI scans showed the precise anatomical geometry of the organ relative to the distal ends of the mandibles (Fig. 3), with XRCT indicating that the organ directly overlies the anterior termination of the YSF stem along the anterior margin of the throat pouch. This configuration provides the mechanical linkages that connect this organ to both bony and soft tissues (that is, the rotation of the bony mandibles⁶ and the expansion of the throat pleats⁷) during the course of a single lunge (Fig. 1).

Following initial observations⁸, it was suggested that the mandibular symphysis in fin whales contained a synovial cavity¹⁹. Our results show that the symphysis is unlike a typical synovial joint with cartilage. Instead, it is bridged by very dense connective tissue, with cruciate fibres, but without true or patent ligaments that originate from the symphyseal groove. In addition, our anatomical and histological analyses indicate that its overall geometry and specialized tissues satisfy the criteria of a sensory organ. Specifically, we argue that the organ responds to localized changes in jaw configuration during lunge feeding, in which the rotation of the bowed jaws will cause deformation of the symphysis, similar to an intervertebral joint²⁰. This action provides the mechanical input to the sensory organ, which the brain receives through the mandibular branch of the trigeminal nerve (Fig. 3). The organ’s asymmetric vascularization and innervation is the first such fluctuating asymmetry reported in the soft tissue of mysticetes, although asymmetric morphology is not unprecedented in toothed

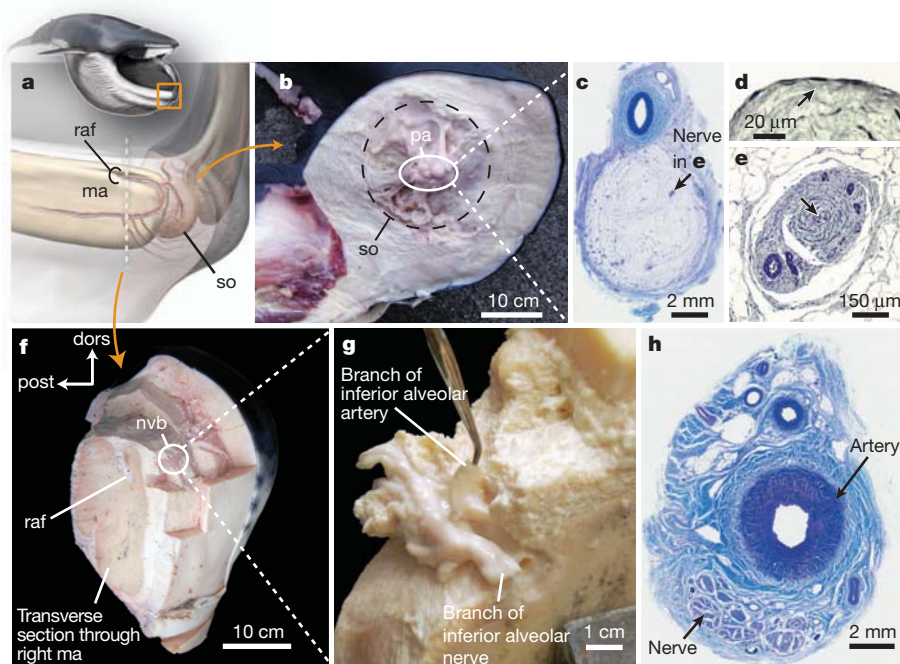


Figure 2 | Anatomy and histology of the rorqual mandibular symphysis, shown in fin whales (*B. physalus*). **a, b**, Mandibular symphysis in sagittal section, showing bony and soft tissue around the sensory organ (so) and papillae (pa) within it. raf, relictu alveolar foramen. **c–e**, Luxol fast blue histological stains showing neurovascular bundles (nvb) and cell layers within

the sensory organ, with arrows pointing to discontinuous squamous cell layer in **d** and nerve termini in **e**. **f, g**, Dissection showing neurovascular bundles in close association with the alveolar groove. dors, dorsal; post, posterior. **h**, Histological section of the tissue shown in **g**. Art by C. Buell.

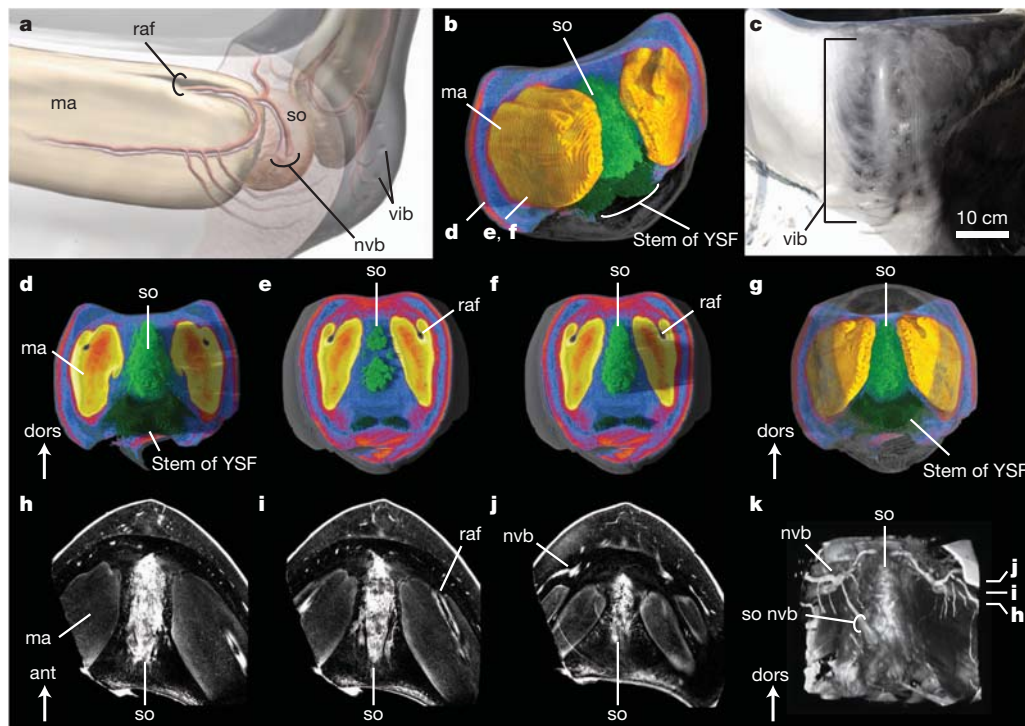


Figure 3 | Digital imaging of the rorqual chin, shown in fin whales (*B. physalus*). **a**, Mandibular symphysis showing bony and neurovascular tissue around the sensory organ (so). **b**, Three-dimensional XRCT rendering of **a**, showing the mandibles (ma), sensory organ and YSF and location of slices **d–f**. **c**, Chin of an adult fin whale. **d–g**, Transverse XRCT slices through the chin

(**g**) showing the sensory organ (light green) in different thicknesses. **h–j**, Coronal MRI slices through the chin. **k**, Asymmetric neurovascular bundles (nvb), with mandibles digitally removed for clarity, and location of MRI slices **j–h**. Ant, anterior. Art by C. Buell.

whales, for both bony tissue (for example, tusks in narwhals (*Monodon monoceros*²¹)) and soft tissue (for example, nasal sacs and bursae²²). The asymmetry associated with this sensory organ may reflect lateralized feeding behaviours (for example, rolling, side gulping) that have individual side biases for several rorqual species^{23,24}.

The new rorqual morphology described here provides yet another example of how aquatic²⁵ and marine mammals²⁶ have evolved unique sensory systems to navigate, communicate and feed in the water²⁷. Mechanoreceptors embedded within fibrous mandibular symphyses have been observed in other terrestrial mammals²⁸, but rorquals alone possess clear mechanoreceptors suspended within a gel-like cavity between a completely unfused symphysis. In terms of lunge-feeding kinematics, anatomical and histological evidence suggest that the organ's mechanosensors assist in controlling and providing neurological information about the configuration of the jaws during rapid lunges. We propose a three-step lunge-feeding model to explain the organ's role during a lunge: (1) using vibrissae present on the external surface of the chin⁹, rorquals register prey fields of sufficient density; (2) the jaws disengage and rotate⁶, thereby compressing and shearing the organ; and (3) the oropharyngeal cavity reaches full expansion, with drag forces acting on the inside of the mouth^{17,18} transmitted to the organ through the YSF⁸ (Fig. 2). According to dynamic modelling studies^{10–13}, rorquals must actively control the rate of mouth opening and throat-pouch expansion to effectively maximize volume captured; we propose that the sensory organ has a key role in coordinating this movement.

The discovery of this organ raises questions about its role in the evolution of lunge feeding in whales. The ancestral mandibular condition of modern cetaceans was probably a fused (or strong sutural) symphysis²⁸ with an elongate surface¹⁴, as seen in nearly all basilosaurids, and a fused symphysis persists in modern toothed whales. By contrast, all living mysticetes possess unfused symphyses^{3,14–16}. Recently described fossil evidence from *Janjucetus hunderi*, a stem 'toothed' mysticete, demonstrates that the transition from fused to unfused

mandibles occurred early in baleen whale evolution, with unfused symphyses in mysticetes having a late Oligocene antiquity (~23–28 Myr ago)¹⁴. Our investigations of the mandibular symphyses in living mysticetes show that bowhead, right and pygmy right whales (*Balaena mysticetus*, *Eubalaena* spp. and *Caperea marginata*, respectively) do not possess a gel-like cavity in their mandibular symphysis, although small papillae intruding into the connective-tissue matrix in this area are patent among adult and immature specimens (see Supplementary Information). The condition of the mandibular symphysis in grey whales (*Eschrichtius robustus*) has only been reported for a decayed neonatal specimen²⁹, although the suggestion of a mucoid centre is broadly similar with the morphology of the organ described here. On the basis of this equivocal evidence, we propose that the organ evolved either at the node of Balaenopteroidea (grey whales and rorquals), or along the stem to crown Balaenopteridae (rorquals)¹⁵. If the former, the organ is a pre-adaptation for lunge feeding; if the latter, the organ evolved in tandem with VGB and specializations of the mandible morphology, such as a laterally deflected coronoid process⁶ and flexible temporomandibular joints⁵, which exhibit a degree of mandible rotation and oropharyngeal cavity expansion far greater than that of grey whales²⁹. Regardless of evolutionary sequence, we argue that multiple lines of evidence indicate that the sensory organ has a key role in lunge feeding by registering the rotation of the mandibles during a lunge and the expansion of the throat pouch through the YSF, all of which evolved before the extremely large body sizes observed in today's rorquals. Despite the opportunity to observe the adult morphology of large rorquals during decades of sustained hunting³⁰, these findings demonstrate how poorly we understand the basic functional morphology of these ecologically important ocean predators.

METHODS SUMMARY

Fieldwork. We recorded macroscopic observations and collected tissue samples from fresh fin whale carcasses that were collected as part of commercial catch operations in Hvalfjörður, Iceland, in 2009 and 2010, as well as carcasses of minke

whales collected by Hrefnuveiðimenn ehf in 2010. For both species, intact carcasses belonged to adult individuals and had been dead for less than 24 h. We directly observed the sensory organ in all of the specimens in which we dissected the mandibular symphysis (see Supplementary Information). All tissue samples from Iceland were transferred and imported to Canada under Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) permits.

Imaging and histology. XRCT scanning was conducted on frozen specimens of *B. physalus* and *B. acutorostrata* at the CT Imaging Centre at the FPInnovations Wood Products Division, using a 4 MeV linear accelerator X-ray source with 0.4 mm spatial resolution, and with a large (1 m) scanning envelope. CT images were rendered using VGStudioMax 2.0 software. MRI scans were conducted on a 1.5 Tesla General Electric superconducting magnetic resonance system at the UBC MRI facility. MRI volume-rendered images were produced at the Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, with a volume-rendering technique using the Inspace application on a Siemens Multimodality Work Place with VE36A software and Fujitsu hardware. Tissues were fixed in standard buffered formalin for histology. Sections were stained using either a generalized haematoxylin and eosin stain or a myelin stain (Luxol fast blue), following protocols devised at Wax-it Histology Services.

Received 27 January; accepted 4 April 2012.

- Sanderson, S. L. & Wassersug, R. in *The Skull, Functional and Evolutionary Mechanisms* Vol. 3 (eds Hanken, J. & Hall, B. K.) 37–112 (Univ. Chicago Press, 1993).
- Friedman, M. *et al.* 100-million-year dynasty of giant planktivorous bony fishes in the Mesozoic seas. *Science* **327**, 990–993 (2010).
- Werth, A. J. in *Feeding: Form, Function and Evolution in Tetrapod Vertebrates* (ed. Schwenk, K.) 475–514 (Academic, 2000).
- Goldbogen, J. A. The ultimate mouthful: lunge feeding in rorqual whales. *Am. Sci.* **98**, 124–131 (2010).
- Brodie, P. F. in *Secondary Adaptation of Tetrapods to Life in Water* (eds Mazin, J.-M. & de Buffrenil, V.) 345–352 (Pfeil, 2001).
- Lambertsen, R., Ulrich, N. & Straley, J. Frontomandibular stay of Balaenopteridae: a mechanism for momentum recapture during feeding. *J. Mamm.* **76**, 877–899 (1995).
- Orton, L. S. & Brodie, P. F. Engulfing mechanics of fin whales. *Can. J. Zool.* **65**, 2898–2907 (1987).
- Pivorunas, A. Fibrocartilage skeleton and related structures of the ventral pouch of balaenopterid whales. *J. Morphol.* **151**, 299–313 (1977).
- Ling, J. K. in *Functional Anatomy of Marine Mammals* (ed. Harrison, R. J.) 387–415 (Academic, 1977).
- Potvin, J., Goldbogen, J. A. & Shadwick, R. E. Passive versus active engulfment: verdict from trajectory simulations of lunge-feeding fin whales. *J. R. Soc. Interface* **6**, 1005–1025 (2009).
- Potvin, J., Goldbogen, J. A. & Shadwick, R. E. Scaling of lunge feeding in rorqual whales: an integrated model of engulfment duration. *J. Theor. Biol.* **267**, 437–453 (2010).
- Goldbogen, J. A. *et al.* Mechanics, hydrodynamics and energetics of blue whale lunge feeding: efficiency dependence on krill density. *J. Exp. Biol.* **214**, 131–146 (2011).
- Goldbogen, J. A. *et al.* Scaling of lunge feeding performance in rorqual whales: mass-specific energy expenditure increases with body size and progressively limits diving capacity. *Funct. Ecol.* **26**, 216–226 (2012).
- Fitzgerald, E. M. G. Archaeocete-like jaws in a baleen whale. *Biol. Lett.* **8**, 94–96 (2012).
- Deméré, T. A., McGowen, M. R., Berta, A. & Gatesy, J. Morphological and molecular evidence for a step-wise evolutionary transition from teeth to baleen in mysticete whales. *Syst. Biol.* **57**, 15–37 (2008).
- Fitzgerald, E. M. G. The morphology and systematics of *Mammalodon colliveri* (Cetacea: Mysticeti), a toothed mysticete from the Oligocene of Australia. *Zool. J. Linn. Soc.* **158**, 367–476 (2010).
- Goldbogen, J. A., Pyenson, N. D. & Shadwick, R. E. Big gulps require high drag for fin whale lunge-feeding. *Mar. Ecol. Prog. Ser.* **349**, 289–301 (2007).
- Lambertsen, R. H. Internal mechanism of rorqual feeding. *J. Mamm.* **64**, 76–88 (1983).
- Brodie, P. F. Noise generated by the jaw actions of feeding fin whales. *Can. J. Zool.* **71**, 2546–2550 (1993).
- Currey, J. D. *Bones: Structure and Mechanics* (Princeton Univ. Press, 2002).
- Best, R. C. The tusk of the narwhal (*Monodon monoceros* L.): interpretation of its function (Mammalia: Cetacea). *Can. J. Zool.* **59**, 2386–2393 (1981).
- McKenna, M. F., Cranford, T. W., Berta, A. & Pyenson, N. D. Morphological diversity of the odontocete melon: implications for acoustic function. *Mar. Mamm. Sci.*, (2011).
- Tershy, B. R. & Wiley, D. M. Asymmetrical pigmentation in the fin whale: a test of two feeding related hypotheses. *Mar. Mamm. Sci.* **8**, 315–318 (1992).
- Canning, C. *et al.* Population-level lateralized feeding behaviour in North Atlantic humpback whales, *Megaptera novaeangliae*. *Anim. Behav.* **82**, 901–909 (2011).
- Catania, K. C., Hare, J. F. & Campbell, K. L. Water shrews detect movement, shape, and smell to find prey underwater. *Proc. Natl Acad. Sci. USA* **105**, 571–576 (2008).
- Czech-Damal, N. U. *et al.* Electroreception in the Guiana dolphin (*Sotalia guianensis*). *Proc. R. Soc. Lond. B* **279**, 663–668 (2012).
- Thewissen, J. G. M. & Nummela, S. *Sensory Evolution on the Threshold: Adaptations in Secondarily Aquatic Vertebrates*. (Univ. California Press, 2008).
- Scapino, R. Morphological investigation into functions of the jaw symphysis in carnivorans. *J. Morphol.* **167**, 339–375 (1981).
- Johnston, C. *et al.* Observations on the musculoskeletal anatomy of the head of a neonate gray whale (*Eschrichtius robustus*). *Mar. Mamm. Sci.* **26**, 186–194 (2010).
- Reeves, R. R. & Smith, T. D. in *Whales, Whaling, and Ocean Ecosystems* (eds Estes, J. A. *et al.*) 82–101 (Univ. California Press, 2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements For logistical support, we thank K. Loftsson and the staff at Hvalur hf; D. Ólafsdóttir, S. D. Halldórsson and G. A. Víkingsson at the Marine Research Institute, Reykjavík; G. Bergmann and the staff at Hrefnuveiðimenn ehf; S. Raverty; P. F. Brodie; and A. Trites. For additional samples and data, we also thank P.-Y. Daoust, G. Williams, the Amarok Hunters and Trappers Organization of Iqaluit, Captain S. Awa and the Inuit whaling crew from Iqaluit, J. Higgins and the Cascadia Research Collective, M. R. Buono, A. van Helden and the Museum of New Zealand Te Papa Tongarewa, and R. E. Fordyce. We also thank T. S. Hunter for assistance with laboratory samples. Comments from D. J. Bohaska, M. T. Carrano, R. B. Irmis, J. G. Mead, J. F. Parham, C. W. Potter and J. Velez-Juarbe improved this manuscript. N.D.P. was supported by a postdoctoral research fellowship from the Natural Sciences and Engineering Research Council of Canada and by funding from the Smithsonian Institution and its Remington Kellogg Fund. J.A.G. was supported by a University Graduate Fellowship for Research from the University of British Columbia, a Scripps Postdoctoral Research Fellowship and NSERC funding to R.E.S.

Author Contributions N.D.P., J.A.G., A.W.V. and R.E.S. conducted field and laboratory investigations and wrote the paper. G.S. provided expertise with XRCT scanning and image reconstruction, and R.L.D. supervised the MRI reconstructions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.D.P. (pyenson@si.edu).

Cryptic peroxisomal targeting via alternative splicing and stop codon read-through in fungi

Johannes Freitag¹, Julia Ast¹ & Michael Böcker¹

Peroxisomes are eukaryotic organelles important for the metabolism of long-chain fatty acids^{1,2}. Here we show that in numerous fungal species, several core enzymes of glycolysis, including glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and 3-phosphoglycerate kinase (PGK), reside in both the cytoplasm and peroxisomes. We detected in these enzymes cryptic type 1 peroxisomal targeting signals (PTS1)³, which are activated by post-transcriptional processes. Notably, the molecular mechanisms that generate the peroxisomal isoforms vary considerably among different species. In the basidiomycete plant pathogen *Ustilago maydis*, peroxisomal targeting of Pfk1 results from ribosomal read-through, whereas alternative splicing generates the PTS1 of Gapdh. In the filamentous ascomycete *Aspergillus nidulans*, peroxisomal targeting of these enzymes is achieved by exactly the opposite mechanisms. We also detected PTS1 motifs in the glycolytic enzymes triose-phosphate isomerase and fructose-bisphosphate aldolase. *U. maydis* mutants lacking the peroxisomal isoforms of Gapdh or Pfk1 showed reduced virulence. In addition, mutational analysis suggests that GAPDH, together with other peroxisomal NADH-dependent dehydrogenases, has a role in redox homeostasis. Owing to its hidden nature, partial peroxisomal targeting of well-studied cytoplasmic enzymes has remained undetected. Thus, we anticipate that further bona fide cytoplasmic proteins exhibit similar dual targeting.

Peroxisomes are highly versatile organelles involved in many biological processes in addition to their common function in fatty-acid metabolism and hydrogen peroxide degradation⁴. In plants, specialized peroxisomes (glyoxysomes) are used during seed germination for breaking down storage lipids⁵. Trypanosomatid parasites contain glycosomes that harbour nearly all glycolytic enzymes, which in all other eukaryotes reside in the cytoplasm⁶. In filamentous ascomycetous fungi, peroxisome-derived Woronin bodies prevent cytoplasmic bleeding in disrupted fungal hyphae and thus allow multicellular development⁷.

Despite the metabolic diversity of peroxisomes, their biogenesis is highly conserved⁸. Peroxisomes receive their membrane from the endoplasmic reticulum and are matured in the cytoplasm^{9,10}. Peroxisomal proteins are not processed during import and pass the membrane in the fully folded state, cofactor bound and even as oligomers¹¹. Most peroxisomal proteins contain a carboxy-terminal PTS1 derived from the prototype tripeptide Ser-Lys-Leu (SKL)³. Further studies revealed that additional C-terminal residues modulate targeting efficiency, with the last 12 C-terminal residues being most important¹². Some peroxisomal proteins contain an alternative signal (PTS2) near the amino terminus¹³. Import of proteins into peroxisomes is mediated by an evolutionary conserved set of proteins known as peroxins (Pex). PTS2-containing proteins are recognized by Pex7, whereas import of PTS1-containing proteins depends on the PTS1 receptor Pex5 (ref. 11).

During a survey of differentially spliced transcripts in the plant pathogenic fungus *U. maydis*, we noticed that the glycolytic enzyme GAPDH (Enzyme Commission (EC) number 1.2.1.12) exists in two isoforms (MIPS *Ustilago maydis* database (MUMDB) accession numbers um02491 and um10167), of which one ends with a

predicted¹⁴ PTS1 motif (-SRL), suggesting peroxisomal localization of this isoform (Gapdh_{pex}) (Fig. 1a). Fusion of the 12 C-terminal amino acids of Gapdh_{pex} to green fluorescent protein (GFP-PTS1_{Gapdh}) resulted in punctate staining (Fig. 1a). These structures were also stained by the peroxisomal marker proteins mCherry-SKL and acyl-coA thioester hydrolase¹⁵ (GFP-Pte1; Supplementary Fig. 1a). We could demonstrate peroxisomal localization of full-length Gapdh_{pex} by expression of GFP-Gapdh_{pex}, derived from the PTS1-containing splice variant (Supplementary Fig. 1b). In peroxisome-deficient mutants ($\Delta pex6$), GFP-Gapdh_{pex} showed diffuse cytoplasmic staining (Supplementary Fig. 1c). To estimate the fraction of peroxisomal Gapdh_{pex}, alternative messenger RNA splice variants were quantified by real-time PCR (rtPCR). Gapdh_{pex} transcripts accounted for about 10% of total *gapdh* mRNA (Fig. 1a and Supplementary Fig. 2). This value is in agreement with the fraction of expressed sequence tags (ESTs) (4 out of 36) listed in the *U. maydis* database corresponding to Gapdh_{pex} (Fig. 1a). Taken together, these data indicate that *U. maydis* expresses a peroxisomal isoform of GAPDH.

The unanticipated peroxisomal localization of a central glycolytic enzyme prompted the screening of other glycolytic enzymes for cryptic PTSs. Analysis of the *pgk1* gene (um04871) encoding PGK (EC 2.7.2.3) revealed that ribosomal read-through¹⁶ of the termination codon results in an extended isoform (Pkg1_{pex}), which contains a functional PTS1 (Fig. 1b). To mimic translational read-through, we replaced the termination codon (TGA) of *pgk1* by a serine codon (TCA). GFP fused to this extended version (GFP-Pkg1_{pex}) was found in peroxisomes (Supplementary Fig. 1d), and this localization was abolished in $\Delta pex6$ mutants (Supplementary Fig. 1c). Expression of C-terminal GFP fusion protein revealed that efficient translational read-through depended on mRNA sequences downstream of the termination codon (Supplementary Fig. 3a). Secondary structure prediction using RNAfold¹⁷ indicated that this region can form a stable stem-loop structure (Supplementary Fig. 3b). To exclude overexpression artefacts GFP-Pkg1 fusion protein was also expressed under control of its endogenous promoter (Supplementary Fig. 3c). Western blot analysis revealed that a considerable amount (20%) of GFP-Pkg1 was expressed as C-terminally extended protein, indicating translational read-through (Fig. 1b and Supplementary Fig. 3d).

We also discovered cryptic peroxisomal targeting of GpdA (GAPDH) and PkgA (PGK) in the ascomycetous fungus *A. nidulans* (Fig. 2). Notably, *A. nidulans* uses reciprocal mechanisms for dual targeting of glycolytic enzymes. Peroxisomal GAPDH is generated by ribosomal read-through (Fig. 2b), whereas the peroxisomal isoform of PGK results from alternative splicing (Fig. 2c). The predicted C-terminal targeting motifs were fused to red fluorescent protein (RFP) and expressed in *Saccharomyces cerevisiae*. Peroxisomal localization of the corresponding fusion proteins (RFP-PTS1_{GpdA} and RFP-PTS1_{PkgA}) demonstrated that both PTS1 signals are functional. Peroxisomal 2,4-dienoyl-CoA reductase encoded by the *S. cerevisiae* *SPS19* gene served as a control¹⁸ (Fig. 2b, c).

Database searches revealed that many fungal genes coding for GAPDH and PGK contain cryptic peroxisomal-targeting motifs

¹Department of Biology, Philipps University Marburg, Karl-von-Frisch-Strasse 8, D-35032 Marburg, Germany.

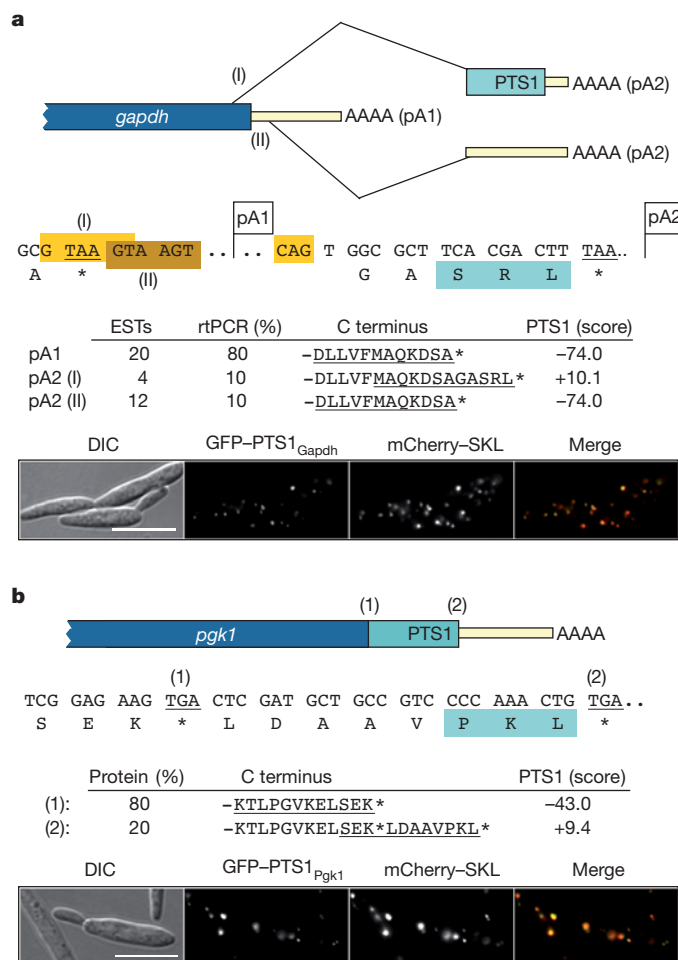


Figure 1 | Cryptic peroxisomal targeting signals in Gapdh and Pgl1 in *U. maydis*. **a**, Schematic drawing of the *gapdh* gene of *U. maydis*. Different transcripts resulting from alternative polyadenylation and splicing are shown. pA1 and pA2 indicate alternative poly(A) sites. (I) and (II) denote alternative 5' splice sites. The C-terminal tripeptide of the peroxisomal isoform is highlighted in blue. Different transcripts were quantified by EST and rtPCR analysis. The PTS1 scores of C-terminal dodecamers (underlined) were obtained using the PTS1-predictor program¹⁴. Positive values indicate high probability of peroxisomal targeting. The C-terminal dodecamer of the splice variant (I) was fused to GFP and co-expressed with mCherry-SKL in *U. maydis*. DIC, differential interference contrast. Scale bar, 10 μ m. **b**, Schematic drawing of the *pgk1* gene of *U. maydis*. Protein sequences resulting from normal termination (1) and translational read-through (2) are indicated. The C-terminal tripeptide is highlighted in blue. The relative fraction of protein was determined by western blot analysis. Peroxisomal-targeting efficiency of the C-terminal dodecamer was determined as described in **a**. Scale bar, 10 μ m.

(Fig. 2a and Supplementary Figs 4 and 5). For some of these genes, ESTs corresponding to predicted peroxisomal isoforms were found in public databases (Supplementary Fig. 6). In addition, a recent study of the *Neurospora crassa* peroxisomal proteome identified GAPDH as a 'PTS-less' peroxisomal protein¹⁹. However, our analysis suggests that *N. crassa* expresses a PTS1-containing isoform of GAPDH by translational read-through (Fig. 2a and Supplementary Fig. 4). We detected hints for peroxisomal targeting of GAPDH and PGK even in the early-diverged zygomycetous fungus *Phycomyces blakesleeanus* (Fig. 2a). The *P. blakesleeanus* *pgk* gene shows hallmarks of translational read-through (Supplementary Fig. 5), whereas three paralogous genes (*gapdh1-3*) encode GAPDH isoforms ending with the tripeptides GAL, GNL and GNA, respectively. Of these isoforms, the PTS1 motif of Gapdh2 was functional in *U. maydis* (Supplementary Fig. 7).

The mosaic distribution of different mechanisms used for dual targeting of GAPDH and PGK in the fungal kingdom suggests several

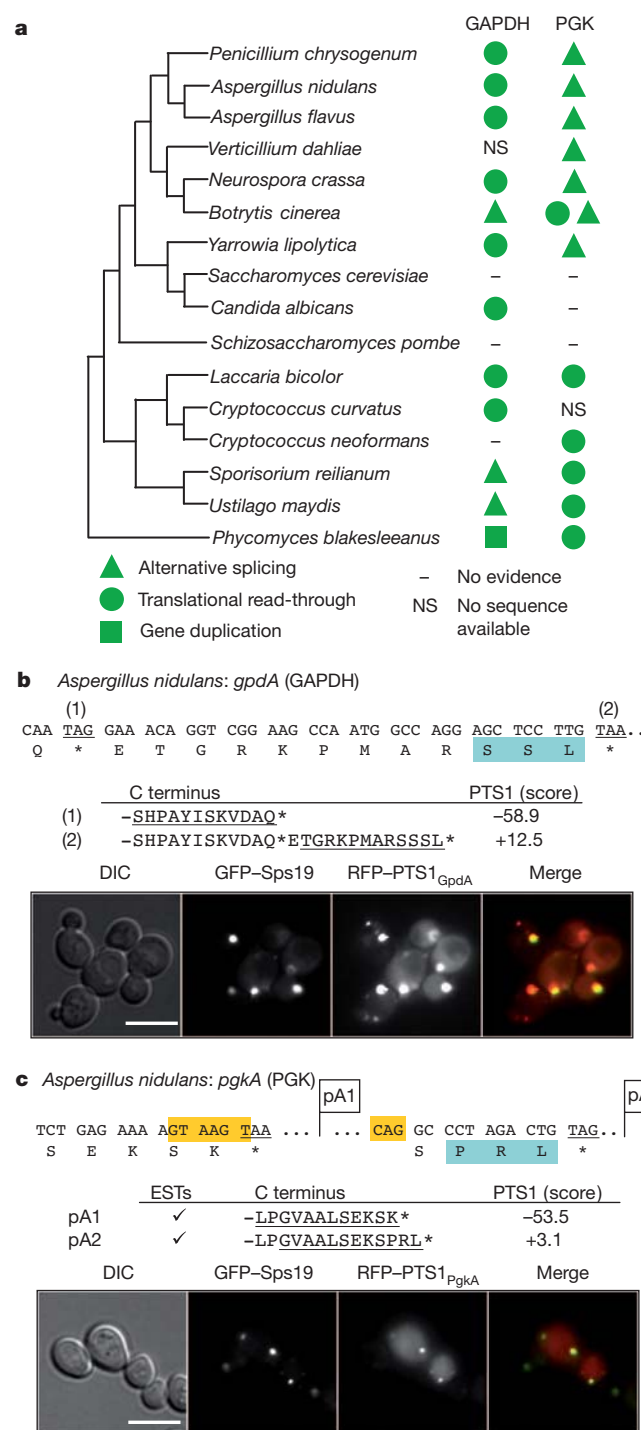


Figure 2 | Dual targeting of GAPDH and PGK occurs in many fungi.

a, Occurrence of dual targeting of glycolytic enzymes in different fungi. Peroxisomal localization of predicted PTS1 motifs was validated by expression of GFP fusion proteins in *S. cerevisiae*. **b**, Protein sequences resulting from normal translational termination and translational read-through of *A. nidulans* *gpdA* (GAPDH) are indicated. The C-terminal tripeptide is highlighted in blue. The PTS1 scores of C-terminal dodecamers (underlined) were obtained using the PTS1-predictor tool¹⁴. The C-terminal dodecamer of sequence (2) was expressed as an RFP fusion in a yeast strain carrying a GFP-Sps19 fusion protein. Scale bar, 5 μ m. **c**, Schematic drawing of the *A. nidulans* *pgkA* gene. Protein sequences resulting from polyadenylation within the intron (pA1) or from splicing (pA2) are indicated. The C-terminal tripeptide is highlighted in blue. ESTs of spliced *pgkA* transcripts were generated by RT-PCR and confirmed by sequencing. Peroxisomal-targeting efficiency of the C-terminal dodecamer was determined as described in **b**. Scale bar, 5 μ m.

evolutionary transitions (Fig. 2a). For PGK a single transition from read-through to splicing may have occurred at the base of the ascomycete branch. The *pgk* gene of the ascomycete *Botrytis cinerea* shows hallmarks of both splicing and read-through, thus corroborating this model (Supplementary Fig. 8). In the case of GAPDH the picture is more complex and at least three independent transitions have to be assumed to explain the phylogenetic distribution of gene duplication, read-through and alternative splicing as mechanisms to achieve dual targeting (Fig. 2a).

To address the biological function of peroxisomal GAPDH and PGK, we deleted the cryptic PTS1 motifs in the respective *U. maydis* genes (Supplementary Fig. 9). Introduction of mutant *gapdh*(Δ PTS1) or *pgk1*(Δ PTS1) into the haploid pathogenic *U. maydis* strain SG20 resulted in a reduction of virulence on maize plants, indicating a role for peroxisomes during pathogenic development (Fig. 3a). This was confirmed by the peroxisome-deficient mutant Δ pex6, which is nearly avirulent and unable to grow on oleate-containing medium (Fig. 3a, b). In addition, Δ pex6 mutants showed a growth defect on glucose (Fig. 3b). A similar phenotype has been described for *pex6* but not for *pex5* mutants in the related basidiomycete *Cryptococcus neoformans*²⁰. This led the authors to speculate that as-yet unidentified PTS2-containing enzymes in peroxisomes may be required for the use of glucose or its metabolites²⁰. However, *C. neoformans*, as well as *U. maydis*, contains two Pex5-related PTS1 receptors²¹, which we termed Pex5a (um02528) and Pex5b (um10172). We could show that in *U. maydis* either Pex5 protein is able to import a PTS1-containing marker protein, and only a *pex5a pex5b* double mutant had a growth defect on glucose (Supplementary Fig. 10). This suggests that PTS1-containing proteins contribute to glucose metabolism in these fungi.

Neither the *U. maydis* *pgk1*(Δ PTS1) nor the *gapdh*(Δ PTS1) mutant showed any growth phenotype on plates containing either oleate or glucose (Supplementary Fig. 9). However, if peroxisomal targeting of other NAD-dependent dehydrogenases was prevented, a synthetic growth defect on glucose and oleate was observed in combination with the *gapdh*(Δ PTS1) mutant (Fig. 3b). PTS1-less mutants of peroxisomal glycerol-3-phosphate dehydrogenase (Gpd1) and malate dehydrogenase (Mdh1) showed either no or only a slight growth defect on glucose (Supplementary Figs 11–13 and Supplementary Table 1). However, introduction of *gapdh*(Δ PTS1) into *mdh1*(Δ PTS1) and *gpd1*(Δ PTS1) mutants considerably increased the growth defect (Supplementary Figs 12, 13 and Supplementary Table 1). If all three dehydrogenases were excluded from peroxisomes, the growth phenotype on glucose and most other sugars was comparable to that of a Δ *pex6* mutant (Fig. 3b, Supplementary Figs 12–14 and Supplementary Table 1). In contrast to Δ *pex6* mutants, the triple mutant showed only slight growth on galactose and was still able to grow, albeit poorly, on oleate as its sole carbon source (Fig. 3b, Supplementary Figs 14, 15 and Supplementary Table 1). This suggests that peroxisomal localization of these dehydrogenases is not critical for long-chain fatty-acid degradation.

The peroxisomal membrane is regarded as impermeable for both ATP and NAD^{22,23}, and thus exchange of redox and ATP equivalents requires either cofactor transport or shuttling of intermediates². Peroxisomal Gpd1 and Mdh1 have both been proposed to be part of the NAD⁺/NADH shuttle systems²³. The synthetic growth phenotype suggests that peroxisomal Gapdh_{pex} also contributes to peroxisomal NAD⁺/NADH homeostasis in *U. maydis*. Recently, a peroxisomal NAD⁺ carrier was described in plants, which contributes to fatty-acid degradation during seed germination²⁵. *U. maydis* contains a putative homologue (um01782), which might support growth on oleate in the absence of the NADH-dependent dehydrogenase shuttle systems.

GPD (EC 1.1.1.8) catalyses the conversion of glycerol-3-phosphate (G3P) into the glycolytic intermediate dihydroxyacetone phosphate (DHAP). During glycolysis, fructose-bisphosphate aldolase (EC 4.1.2.13; FBA) generates DHAP and glyceraldehyde-3-phosphate (GAP) by breakdown of fructose-1,6-bisphosphate. Triose-phosphate isomerase

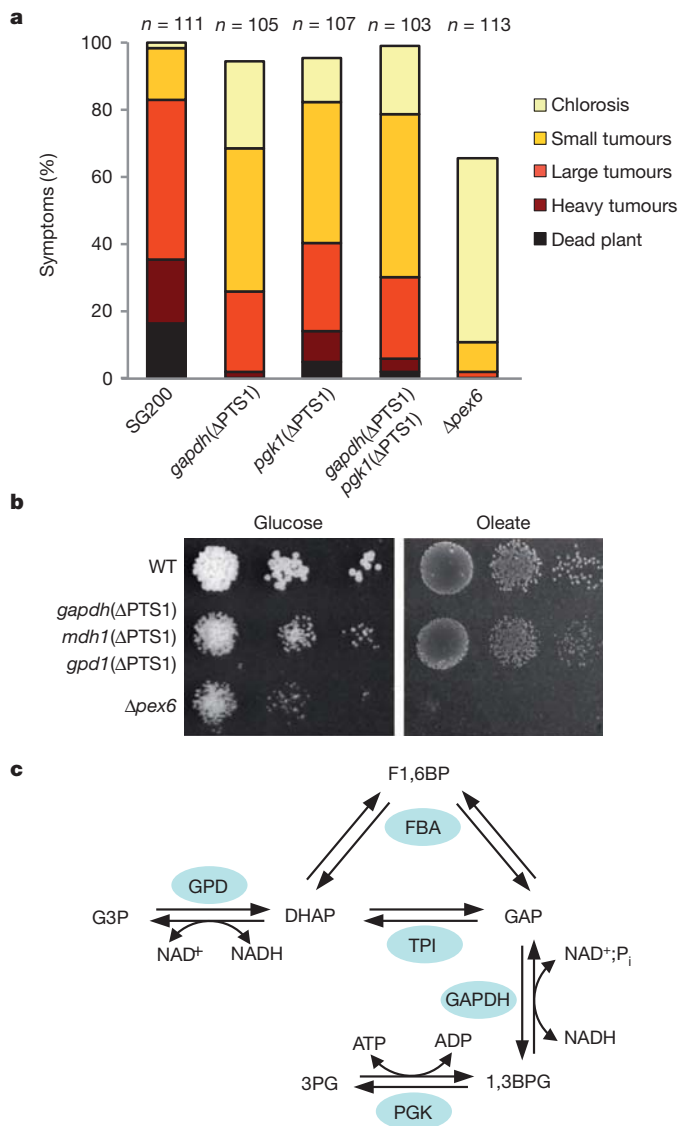


Figure 3 | Peroxisomal localization of glycolytic enzymes is important for virulence and growth in *U. maydis*. **a**, The wild-type strain SG200 and its mutant derivatives *gapdh*(Δ PTS1), *pgk1*(Δ PTS1), *gapdh*(Δ PTS1) *pgk1*(Δ PTS1) and Δ *perx6* were used to infect 7-day-old maize seedlings. Disease symptoms were scored 10 days post infection. Symptoms were grouped into the categories depicted on the right. Average values of three independent experiments are given as percentage of the total number (*n*) of infected plants. **b**, Serial dilutions of indicated strains were spotted on glucose- and oleate-containing solid medium and photographed after 48 h at 28 °C. WT, wild type. **c**, Schematic illustration of reactions catalysed in peroxisomes by glycolytic/gluconeogenic enzymes. This metabolic network may function in redox/ATP shuttling or as buffer system to cope with perturbations of redox/ATP equivalents. 1,3BPG, 1,3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; F1,6BP, fructose-1,6-bisphosphate; G3P, glycerol-3-phosphate.

(EC 5.3.1.1; TPI) is responsible for rapid isomerization of DHAP and GAP (Fig. 3c). We noticed that *U. maydis* Tpi1 also contains a cryptic PTS1 motif, which is activated by translational read-through (Supplementary Fig. 16). Although full-length Fba1 seems to localize in the cytoplasm, its protein sequence contains a functional PTS1 (Supplementary Fig. 16). Therefore we assume that this motif has only low efficiency and targets a fraction of Fba1 to peroxisomes. Peroxisomal localization of TPI and FBA has been observed in *Penicillium chrysogenum*²⁶. Here, we could also show that in *A. nidulans* both TPI and FBA contain functional PTS1 motifs suggesting at least partial peroxisomal localization (Supplementary Fig. 17). Recently, it was shown that in *Candida albicans* and *Yarrowia lipolytica* alternative

splicing directs dual localization of 6-phosphogluconate dehydrogenase and Mdh1, respectively^{27,28}.

Our results demonstrate that in many fungal species peroxisomes contain a metabolic network, which was previously thought to reside only in the cytoplasm. All reactions of this network are reversible and thus can work both in the glycolytic and the gluconeogenic direction. Because some reactions depend on NAD⁺/NADH and ATP/ADP, the whole network can act as both sink and source for these cofactors. Therefore, this system could act as a buffer to compensate perturbations of NAD⁺/NADH and ATP/ADP ratios. Alternatively, this peroxisomal network may establish a separate pool of ATP to deal with the inherent risk of the 'turbo' design of glycolysis²⁹, as has been shown for glycosomes of trypanosomes³⁰. Compartmentalization may also prevent futile cycles of ATP hydrolysis resulting from antagonistic reactions of glycolysis and gluconeogenesis.

Here we have demonstrated that fungi use a variety of mechanisms (ribosomal read-through, differential splicing and low-efficiency PTS1 signals) to achieve partial peroxisomal targeting of glycolytic enzymes. Although we could not detect dual localization of glycolytic enzymes in plants, mammalian PGK proteins contain functional PTS1 motifs (Supplementary Fig. 18). Thus, it will be of particular interest to investigate other cytoplasmic proteins for dual localization. Such analysis could add more unanticipated enzymes to the peroxisomal proteome and will contribute to a more comprehensive understanding of the metabolic capacity of peroxisomes.

METHODS SUMMARY

All experiments were performed in accordance with established protocols. For *U. maydis* strains used, see Supplementary Table 2. Details of transformation, plant infections, microscopy, molecular biology and computational analysis are provided in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 October 2011; accepted 21 March 2012.

- De Duve, C. & Baudhuin, P. Peroxisomes (microbodies and related particles). *Physiol. Rev.* **46**, 323–357 (1966).
- Poirier, Y., Antonenkov, V. D., Glumoff, T. & Hiltunen, J. K. Peroxisomal β -oxidation – a metabolic pathway with multiple functions. *Biochim. Biophys. Acta* **1763**, 1413–1426 (2006).
- Gould, S. J., Keller, G. A., Hosken, N., Wilkinson, J. & Subramani, S. A conserved tripeptide sorts proteins to peroxisomes. *J. Cell Biol.* **108**, 1657–1664 (1989).
- Gabaldon, T. Peroxisome diversity and evolution. *Phil. Trans. R. Soc. Lond. B* **365**, 765–773 (2010).
- Graham, I. A. Seed storage oil mobilization. *Annu. Rev. Plant Biol.* **59**, 115–142 (2008).
- Michels, P. A., Bringaud, F., Herman, M. & Hannaert, V. Metabolic functions of glycosomes in trypanosomatids. *Biochim. Biophys. Acta* **1763**, 1463–1477 (2006).
- Jedd, G. Fungal evo-devo: organelles and multicellular complexity. *Trends Cell Biol.* **21**, 12–19 (2011).
- Purdue, P. E. & Lazarow, P. B. Peroxisome biogenesis. *Annu. Rev. Cell Dev. Biol.* **17**, 701–752 (2001).
- Hoepfner, D., Schildknecht, D., Braakman, I., Philippsen, P. & Tabak, H. F. Contribution of the endoplasmic reticulum to peroxisome formation. *Cell* **122**, 85–95 (2005).
- Tabak, H. F. *et al.* Formation of peroxisomes: present and past. *Biochim. Biophys. Acta* **1763**, 1647–1654 (2006).
- Girzalsky, W., Saffian, D. & Erdmann, R. Peroxisomal protein translocation. *Biochim. Biophys. Acta* **1803**, 724–731 (2010).
- Brocard, C. & Hartig, A. Peroxisome targeting signal 1: is it really a simple tripeptide? *Biochim. Biophys. Acta* **1763**, 1565–1573 (2006).
- Lazarow, P. B. The import receptor Pex7p and the PTS2 targeting sequence. *Biochim. Biophys. Acta* **1763**, 1599–1604 (2006).
- Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A. & Eisenhaber, F. Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. *J. Mol. Biol.* **328**, 581–592 (2003).
- Jones, J. M., Nau, K., Geraghty, M. T., Erdmann, R. & Gould, S. J. Identification of peroxisomal acyl-CoA thioesterases in yeast and humans. *J. Biol. Chem.* **274**, 9216–9223 (1999).
- Baranov, P. V., Gesteland, R. F. & Atkins, J. F. Recoding: translational bifurcations in gene expression. *Gene* **286**, 187–201 (2002).
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).
- Gurvitz, A. *et al.* The *Saccharomyces cerevisiae* peroxisomal 2,4-dienoyl-CoA reductase is encoded by the oleate-inducible gene *SPS19*. *J. Biol. Chem.* **272**, 22140–22147 (1997).
- Managadze, D. *et al.* A proteomic approach towards the identification of the matrix protein content of the two types of microbodies in *Neurospora crassa*. *Proteomics* **10**, 3222–3234 (2010).
- Idnurm, A., Giles, S. S., Perfect, J. R. & Heitman, J. Peroxisome function regulates growth on glucose in the basidiomycete fungus *Cryptococcus neoformans*. *Eukaryot. Cell* **6**, 60–72 (2007).
- Kiel, J. A., Veenhuis, M. & van der Kleij, I. J. *PEX* genes in fungal genomes: common, rare or redundant. *Traffic* **7**, 1291–1303 (2006).
- van Roermund, C. W., Elgersma, Y., Singh, N., Wanders, R. J. & Tabak, H. F. The membrane of peroxisomes in *Saccharomyces cerevisiae* is impermeable to NAD(H) and acetyl-CoA under *in vivo* conditions. *EMBO J.* **14**, 3480–3486 (1995).
- Visser, W. F., van Roermund, C. W., Ijlst, L., Waterham, H. R. & Wanders, R. J. Metabolite transport across the peroxisomal membrane. *Biochem. J.* **401**, 365–375 (2007).
- Rottensteiner, H. & Theodoulou, F. L. The ins and outs of peroxisomes: co-ordination of membrane transport and peroxisomal metabolism. *Biochim. Biophys. Acta* **1763**, 1527–1540 (2006).
- Bernhardt, K., Wilkinson, S., Weber, A. P. & Linka, N. A peroxisomal carrier delivers NAD and contributes to optimal fatty acid degradation during storage oil mobilization. *Plant J.* **69**, 1–13 (2011).
- Kiel, J. A. *et al.* Matching the proteome to the genome: the microbody of penicillin-producing *Penicillium chrysogenum* cells. *Funct. Integr. Genomics* **9**, 167–184 (2009).
- Strijbis, K., van den Burg, J. F., Visser, W., van den Berg, M. & Distel, B. Alternative splicing directs dual localization of *Candida albicans* 6-phosphogluconate dehydrogenase to cytosol and peroxisomes. *FEMS Yeast Res.* **12**, 61–68 (2012).
- Kabran, P., Rossignol, T., Gaillardin, C., Nicaud, J. M. & Neuveglise, C. Alternative splicing regulates targeting of malate dehydrogenase in *Yarrowia lipolytica*. *DNA Res.* doi:10.1093/dnares/dss007 (24 February 2012).
- Teusink, B., Walsh, M. C., van Dam, K. & Westerhoff, H. V. The danger of metabolic pathways with turbo design. *Trends Biochem. Sci.* **23**, 162–169 (1998).
- Haanstra, J. R. *et al.* Compartmentation prevents a lethal turbo-explosion of glycolysis in trypanosomes. *Proc. Natl Acad. Sci. USA* **105**, 17718–17723 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to J. Heitman and W. K. Holloman for reading of the manuscript. We thank R. Fischer and G. Barth for providing strains and DNA. We acknowledge technical assistance by M. Piscator. We thank D. Lanver and T. Stehlik for discussions. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG-GK1216). J.A. received a fellowship from the Marburg University Research Academy (MARA).

Author Contributions M.B. and J.F. designed this study. J.F. and J.A. performed the experiments. All authors contributed to data analysis. M.B. and J.F. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.B. (boelker@uni-marburg.de).

METHODS

Strains and growth conditions. The *Escherichia coli* strain Top Ten (Invitrogen) was used for all cloning procedures and amplification of plasmids³¹. All *U. maydis* strains used in this study are listed in Supplementary Table 2. Liquid cultures of *U. maydis* were grown in either yeast-extract peptone dextrose (YEPD)³⁴ or yeast nitrogen base (YNB; Difco) supplemented with 2% glucose at 28 °C. Solid media were based on YEPD, YNB or potato dextrose (PD; Difco). Special supplements are indicated in respective figures. Growth assays in liquid culture were performed using YNB supplemented with either 2% glucose or 0.3% oleate/0.1% Tween 40 at 28 °C. The *S. cerevisiae* strain S288c (Euroscarf) was used for yeast manipulations. *S. cerevisiae* cells were grown on synthetic complete (SC) or SC-Leu medium at 30 °C.

Transformation. *U. maydis* was transformed as described previously³². To achieve comparable expression levels, respective plasmid constructs were integrated in the *cbx* locus³³. All deletion mutants were verified using Southern blot analysis³¹. At least two independent mutants were analysed for phenotypes. *S. cerevisiae* was transformed using standard procedures³⁴.

Generation of plasmids. Deletion constructs for *U. maydis* were generated using a PCR-based approach and a SfiI-based cloning strategy³⁵. Ectopic expression of GFP or mCherry fusion proteins was reached by using the constitutive *otf* promoter^{36,37}. For detection of peroxisomes in *S. cerevisiae*, the *SPS19* gene was expressed as a GFP fusion using the plasmid pYM-N9 (ref. 38). Predicted PTS1 signals were tested for peroxisomal-targeting efficiency by fusion to tagRFP expressed under control of the *S. cerevisiae* *PGK1* promoter in the plasmid pCT-320 (gift from C. Taxis). All primer and plasmid sequences are available on request.

Microscopy. Cells from logarithmically growing cultures were placed on agarose cushions and visualized by differential interference contrast and epifluorescence microscopy using a Zeiss Axiophot 200 microscope (Göttingen, Germany). Images were taken using a cooled CCD camera (Hamamatsu Orca-ER, Herrsching, Germany) with an exposure time of 50–300 ms. Images were processed using ImageJ (National Institutes of Health).

Nucleic acid procedures and rtPCR. All DNA procedures were done as described³¹. Genomic DNA from *S. cerevisiae* and *U. maydis* was isolated as described³⁹. For RNA isolation, the *U. maydis* strain Bub8 was grown to a density of 8×10^6 cells ml⁻¹. Cells were collected and ground in liquid nitrogen. RNA was extracted with Trizol (Invitrogen) and purified using RNeasy kit (Qiagen). For rtPCR complementary DNA was synthesized using the First Strand cDNA Kit (Fermentas) with 1 µg total RNA. rtPCR was performed using the MAXIMA SYBR-Green qPCR Master-Mix (Fermentas) on a Biorad iCycler. Reaction conditions were as follows: 5 min 95 °C followed by 45 cycles of 15 s 95 °C/15 s 60 °C/30 s 72 °C. Primer sequences are shown in Supplementary Fig. 2 and are available on request. 3'RACE was performed using 100 ng of *A. nidulans* cDNA (gift from R. Fischer). Reactions were performed using Phusion DNA polymerase (New England Biolabs). Reaction conditions were as follows: 5 min 95 °C followed by 32 cycles of 15 s 95 °C/15 s 55 °C/30 s 72 °C and the resulting products were subcloned and sequenced. Northern blotting was performed essentially as described³¹.

Protein isolation and western blotting. For protein isolation the *U. maydis* strain Bub8 was grown to a density of 8×10^6 cells ml⁻¹. Cells were collected and ground in liquid nitrogen. Pellets were resuspended in 0.1 M Tris-buffered saline supplemented with 0.1% Tween 20 and cells were destroyed mechanically on a Vibrax shaker. Proteins were separated using SDS-PAGE and blotted on polyvinylidene difluoride membranes. Proteins were detected with antibodies against GFP (Santa Cruz Biotechnology) and tubulin (Calbiochem).

Plant infections. *U. maydis* was grown in YEPSL medium (0.4% yeast extract, 0.4% peptone, 2% sucrose) to a density of 8×10^6 cells ml⁻¹ and concentrated in H₂O to a final density of 1×10^7 cells ml⁻¹. The suspension was used to infect 7-day-old seedlings of Early Golden Bantam (Olds Seeds) by injecting 0.5 ml into each seedling. Ten days after infection, disease symptoms were evaluated according to standard rating criteria⁴⁰. All experiments were done in three biological replicates.

Computational analysis. The following databases were used for gene comparisons, EST analysis and to search for cryptic PTSs: MUMDB: <http://mips.helmholtz-muenchen.de/genre/proj/ustilago/>; NCBI: <http://www.ncbi.nlm.nih.gov/>; *Saccharomyces* Genome Database (SGD): <http://www.yeastgenome.org/>; Joint Genome Institute (JGI): <http://www.jgi.doe.gov/>; and Broad Institute: <http://www.broadinstitute.org/>. Phylogenetic analysis was performed using ClustalX (<http://www.clustal.org/>) with ribosomal DNA sequences downloaded from Silva (<http://www.arb-silva.de/>). RNA secondary structure was predicted using RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>). PTS1 predictions were performed using the general function or the metazoan-specific function of PTS1 Predictor (<http://mendel.imp.ac.at/jspcgi/cgi-bin/pts1/pts1.cgi>).

1. Sambrook, J. & Russell, D. *Molecular Cloning: a Laboratory Manual* 3rd edn (Cold Spring Harbor Laboratory Press, 2001).
2. Schulz, B. *et al.* The alleles of *U. maydis*, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif. *Cell* **60**, 295–306 (1990).
3. Loubradou, G., Brachmann, A., Feldbrügge, M. & Kahmann, R. A homologue of the transcriptional repressor Ssn6p antagonizes cAMP signalling in *Ustilago maydis*. *Mol. Microbiol.* **40**, 719–730 (2001).
4. Guthrie, C. & Fink, G. R. *Guide to Yeast Genetics and Molecular and Cell Biology*. (Academic, 2002).
5. Brachmann, A., König, J., Julius, C. & Feldbrügge, M. A reverse genetic approach for generating gene replacement mutants in *Ustilago maydis*. *Mol. Genet. Genomics* **272**, 488 (2004).
6. Schink, K. O. & Böcker, M. Coordination of cytokinesis and cell separation by endosomal targeting of a Cdc42-specific guanine nucleotide exchange factor in *Ustilago maydis*. *Mol. Biol. Cell* **20**, 1081–1088 (2009).
7. Spellig, T., Bottin, A. & Kahmann, R. Green fluorescent protein (GFP) as a new vital marker in the phytopathogenic fungus *Ustilago maydis*. *Mol. Gen. Genet.* **252**, 503–509 (1996).
8. Janke, C. *et al.* A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–962 (2004).
9. Hoffman, C. S. & Winston, F. A ten minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation in *Escherichia coli*. *Gene* **57**, 267–272 (1987).
10. Kämper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97–101 (2006).

CORRECTIONS & AMENDMENTS

ERRATUM

doi:10.1038/nature11138

Erratum: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability

Ben B. B. Booth, Nick J. Dunstone, Paul R. Halloran,
Timothy Andrews & Nicolas Bellouin

Nature **484**, 228–232 (2012)

The accepted date should read 08 February 2012, rather than 2011.
This error has been corrected in the PDF and HTML versions online.

Evolution of the chalcone–isomerase fold from fatty–acid binding to stereospecific catalysis

Micheline N. Ngaki^{1*}, Gordon V. Louie^{2*}, Ryan N. Philippe^{2*}, Gerard Manning³, Florence Pojer², Marianne E. Bowman², Ling Li¹, Elise Larsen², Eve Syrkin Wurtele¹ & Joseph P. Noel²

Specialized metabolic enzymes biosynthesize chemicals of ecological importance, often sharing a pedigree with primary metabolic enzymes¹. However, the lineage of the enzyme chalcone isomerase (CHI) remained unknown. In vascular plants, CHI-catalysed conversion of chalcones to chiral (*S*)-flavanones is a committed step in the production of plant flavonoids, compounds that contribute to attraction, defence² and development³. CHI operates near the diffusion limit with stereospecific control^{4,5}. Although associated primarily with plants, the CHI fold occurs in several other eukaryotic lineages and in some bacteria. Here we report crystal structures, ligand-binding properties and *in vivo* functional characterization of a non-catalytic CHI-fold family from plants. *Arabidopsis thaliana* contains five actively transcribed genes encoding CHI-fold proteins, three of which additionally encode amino-terminal chloroplast-transit sequences. These three CHI-fold proteins localize to plastids, the site of *de novo* fatty-acid biosynthesis in plant cells. Furthermore, their expression profiles correlate with those of core fatty-acid biosynthetic enzymes, with maximal expression occurring in seeds and coinciding with increased fatty-acid storage in the developing embryo. *In vitro*, these proteins are fatty-acid-binding proteins (FAPs). FAP knockout *A. thaliana* plants show elevated α -linolenic acid levels and marked reproductive defects, including aberrant seed formation. Notably, the FAP discovery defines the adaptive evolution of a stereospecific and catalytically ‘perfected’ enzyme⁶ from a non-enzymatic ancestor over a defined period of plant evolution.

CHI (EC 5.5.1.6) catalyses the intramolecular and stereospecific cyclization of chalcones to chiral flavanones by a Michael addition reaction⁷. The origin of CHI has been a mystery, because of the apparent absence of a related protein from primary metabolism⁸. CHI-like homologues in fungi and bacteria⁹ lack both key catalytic residues and the chalcone-binding site of bona fide CHI (Fig. 1a–c and Supplementary Fig. 1). Phylogenetic analysis shows that CHI is restricted to vascular plants, and is derived from FAP3. FAP3 forms one of two branches of FAPs (FAP1 and FAP2 form the other) found in derived and basal plants and is probably homologous to CHI-fold proteins from protists and fungi (Supplementary Figs 2 and 3 and Supplementary Files 1 and 2). *A. thaliana* has five genes (and one pseudogene not shown) encoding CHI-fold proteins: two CHI/CHI-like (CHIL) members, *AtCHI* (At3g55120) and *AtCHIL* (At5g05270), and three FAP members, *AtFAP1*, *AtFAP2* and *AtFAP3* (At3g63170, At2g26310 and At1g53520, respectively). Pairwise sequence identities of the shared CHI-fold domain range from 10 to 63%. *AtCHI* is a bona fide enzyme, and mutations of its gene result in plants devoid of flavonoids¹⁰. The active site of *AtCHI* retains residues important for catalytic activity, and *AtCHI* catalyses the *in vitro* formation of (2*S*)-naringenin (Fig. 1c). In contrast, *AtCHIL* has substitutions of several catalytic residues (Supplementary Fig. 3). Likewise, *AtFAP1*, *AtFAP2* and *AtFAP3* bear substitutions at nearly all of the critical catalytic positions of *AtCHI*, and are devoid of catalytic activity

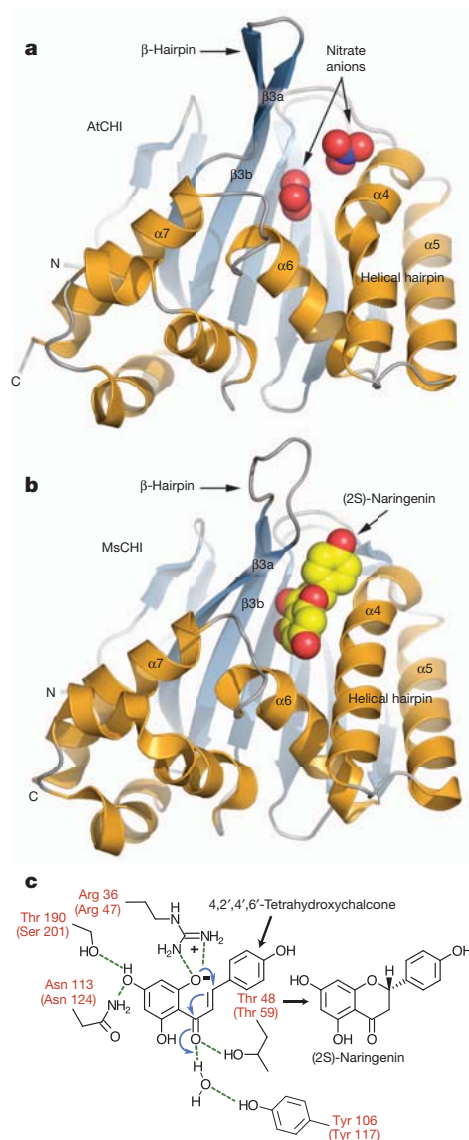


Figure 1 | CHI fold and catalytic reaction. **a**, Ribbon diagram²¹ of the *AtCHI* X-ray crystal structure, colour-coded and labelled according to Jez *et al.*⁸. Two nitrate anions associate with catalytic residues in the substrate-binding site. **b**, Ribbon diagram²¹ of the *MsCHI* structure bound to (2*S*)-naringenin⁸. **c**, Chalcone is converted to (2*S*)-flavanone (for example, naringenin) using a combination of electrostatic catalysis and water-mediated charge stabilization during a stereospecific Michael-type addition reaction^{5,7,8}. Residue numbers for *AtCHI* appear in parentheses. Catalytic residues are coloured red.

¹Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, Iowa 50011, USA. ²Howard Hughes Medical Institute, Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. ³Razavi Newman Center for Bioinformatics, The Salk Institute for Biological Studies, La Jolla, California 92037, USA.

*These authors contributed equally to this work.

(Supplementary Fig. 3). Studies with green fluorescent protein (GFP) fused to full-length FAP1, FAP2 and FAP3 indicate that these proteins localize in the plastid stroma (Supplementary Fig. 4a–d).

We determined X-ray crystal structures of AtCHI, AtCHIL and the CHI-like domains of AtFAP1 and AtFAP3, and compared these with the previously reported structure of MsCHI⁸. Despite low sequence conservation, all structures align within 2.3 Å root-mean-squared deviation for the backbone atoms of equivalent residues (Supplementary Fig. 5). Ligand-binding pockets reside within the helical layer, bounded by two α -helical segments (helix–turn–helix $\alpha 6$ – $\alpha 7$ and helical hairpin $\alpha 4$ – $\alpha 5$), β -hairpin $\beta 3a$ – $\beta 3b$ and the face of the core β -sheet, $\beta 3a$ – $\beta 3f$. Notably, the largest differences in backbone conformations of MsCHI, AtCHI and AtCHIL compared with AtFAP1 and AtFAP3 are confined to the secondary structure elements surrounding the ligand-binding pockets (Supplementary Fig. 5).

The most intriguing finding from the crystallographic structures is that the ligand-binding pockets of the recombinant FAPs are occupied by fatty acids (Fig. 2a–d). Clearly evident in electron-density maps are two abutting molecules of lauric acid (C12:0) in AtFAP1 (Fig. 2a, b), and one molecule of palmitic acid (C16:0) in AtFAP3 (Fig. 2c, d). The ligand-binding site of the FAPs encompasses a largely buried, non-polar cavity that sequesters the aliphatic chain(s) of the fatty acids, and a conserved Arg–Tyr pair that tethers a fatty-acid carboxylate group (Fig. 2b, d and Supplementary Fig. 6). By comparison, CHI and CHIL possess more sterically restricted pockets owing to the presence of larger aliphatic residues and an inward shift of the secondary structure elements surrounding their ligand-binding clefts (Figs 1 and 2 and Supplementary Fig. 5). The CHI substrate-binding pocket also includes several polar residues that form key hydrogen-bond interactions with the chalcone substrate (Fig. 1c). The corresponding residues in the FAPs are non-polar. AtCHIL possesses a distinct conservation pattern of the active site pocket (Supplementary Fig. 3). Extraction of ligands with ethanol

followed by high-performance liquid chromatography–mass spectrometry (HPLC–MS) analyses confirms that all three FAPs (including AtFAP2, which was recalcitrant to crystallization) associate with saturated fatty acids (Fig. 2e). In contrast, with AtCHI and AtCHIL produced under similar conditions, no associated fatty acids were detected analytically or crystallographically. Exogenous fatty acids were not supplied during FAP expression in *Escherichia coli*, purification or crystallization, so the ligands bound to the FAPs were probably acquired during expression and retained during purification to homogeneity. Indeed, these saturated fatty acids are among the most abundant fatty acids in *E. coli*¹¹.

There are indications that the true ligands of the FAPs *in planta* may be different fatty acids. In AtFAP1, the presence of two distinct non-polar tunnels, each occupied by the aliphatic chain of a C12:0 molecule (Fig. 2b and Supplementary Fig. 6a), may indicate a proclivity for longer chain fatty-acid recognition. In AtFAP3, the horseshoe-shaped conformation of C16:0 portends a preference for binding a *cis*-unsaturated C16 or C18 FA, which typically assumes a bent conformation (Fig. 2d and Supplementary Fig. 6b). Notably, AtFAP1 sequesters exogenously supplied α -linolenic acid (C18:3), which presumably displaces the bound fatty acids acquired during expression in *E. coli* (Fig. 2e). We further assessed the binding by the FAPs of a series of common fatty acids using thermal-stability assays¹² (Supplementary Table 1). AtFAP3 interacts with most fatty acids tested and has maximal relative affinity for C16:0, though C18:0 induces greater thermal stabilization (Supplementary Fig. 7). AtFAP1 preferentially interacts with saturated fatty acids. The strongest stabilizing effect was observed with C18:3 (Supplementary Fig. 8), consistent with the displacement analyses (Fig. 2e) and perhaps a function of AtFAP1 in C18:3 metabolism as indicated also by phenotypic analysis. The thermal stability of AtCHI, included in these analyses as a negative control, was unaffected by the addition of fatty acids (Supplementary Fig. 9).

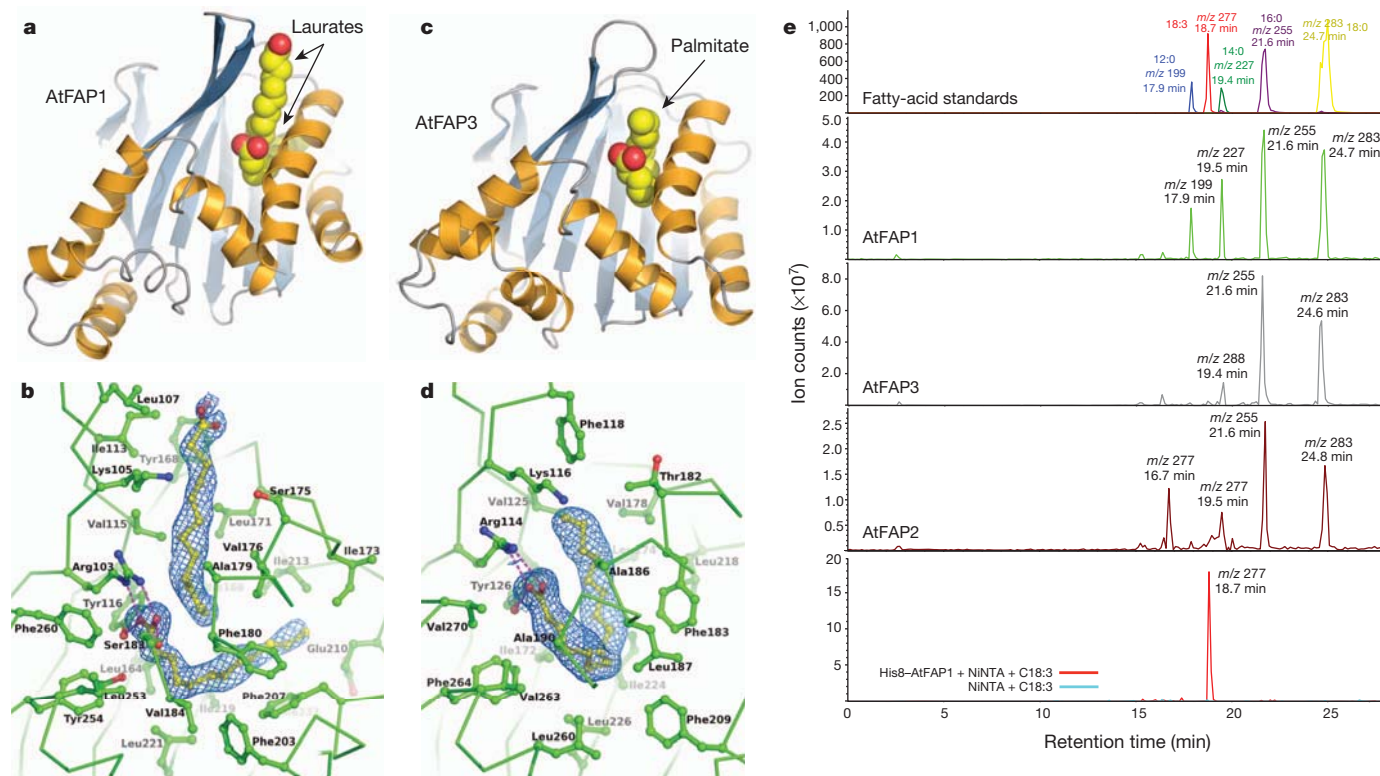


Figure 2 | Three-dimensional structure and ligand binding of FAPs. a, Ribbon diagram²¹ of AtFAP1 oriented and colour-coded as in Fig. 1. The bound C12:0 molecules are shown as van der Waals spheres where carbon is yellow and oxygen is red. b, Close-up view of the AtFAP1 fatty-acid binding sites²¹ with the experimental electron density of each C12:0 shown at 1 σ for a SIGMAA-weighted $2F_o - F_c$ map²². c, Ribbon diagram²¹ of AtFAP3 rendered

as in a. d, Close-up view of the AtFAP3 fatty-acid binding site²¹ with the C16:0 ligand shown as in b. e, Analysis of ligands associated with purified AtCHI-fold proteins separated and detected by reversed-phase HPLC–MS. The y-axes represent negative-ion counts for selected masses of anionic forms of fatty acids. The bottom panel depicts binding of C18:3 by His8-tagged AtFAP1 coupled to Ni²⁺-affinity resin (red) and by the control (blue).

We have obtained, through a variety of approaches, considerable evidence supporting the involvement of *AtFAPs* in fatty-acid metabolism. Network analysis of transcriptomic data^{13,14} demonstrated maximal expression of *AtFAP1* and *AtFAP3* in seeds at 6 days after flowering (Supplementary Fig. 10a), coinciding with the accumulation of storage lipids in the developing embryo¹⁵. *AtFAP1* and *AtFAP3* co-express with genes encoding fatty-acid biosynthetic enzymes (Supplementary Fig. 10b). Developmental patterns of *AtFAP1*, *AtFAP2* and *AtFAP3* expression were characterized in *A. thaliana* lines expressing a β -glucuronidase (GUS) reporter under control of the respective promoters (Supplementary Figs 11, 12 and 13). The FAPs are highly expressed in developing cotyledons, young seedlings, roots, seeds, embryos, macrospores, preanthesis and tapetum. However, although *AtFAP2* is expressed throughout the life of the plant, expression of *AtFAP1* and especially *AtFAP3* is restricted to developing and reproductive tissues.

Homozygous *Atfap1*- and *Atfap2*-null plants were propagated from transfer DNA (T-DNA) insertion lines (Supplementary Figs 14 and 15). These *Atfap1* and *Atfap2* null plants are indistinguishable from wild-type (WT) plants during vegetative growth (Supplementary Figs 17 and 18), but the *Atfap1* mutants show marked differences during reproductive stages. Specifically, *Atfap1* siliques are shorter than those of WT (~18 mm versus ~20 mm) (Fig. 3a), mutant siliques contain 10–20% abnormal ovules (Fig. 3b), and the number of seeds per silique and the yield of viable seeds per plant are reduced in the *Atfap1* null lines (Fig. 3c). The *AtFAP3*-RNAi homozygote lines also present these

phenotypic alterations. In addition, the vegetative phenotype of *AtFAP3*-RNAi is altered, demonstrating early bolting, fast growth, increased branching, reduced apical dominance and reduced overall plant height (Supplementary Fig. 19). No *AtFAP3*-RNAi line with more than an approximately 50% reduction in *AtFAP3* RNA (relative to WT) could be identified, nor are any T-DNA tagged *Atfap3* knockout lines available. These data thus indicate that a functional *AtFAP3* is required for normal plant and embryo development. Double mutants bearing pairwise combinations of homozygous *Atfap1*, *Atfap2*, and *AtFAP3*-RNAi all show additive developmental effects (Supplementary Fig. 18).

Reciprocal crosses between WT and *Atfap1* or *AtFAP3*-RNAi lines point to the maternal inheritance of the aberrant reproductive phenotypes, as an elevated rate of ovule abortion is observed only in crosses in which the maternal parent is a null mutant (Supplementary Figs 20–21, Supplementary Tables 2–7). The morphological phenotypes of more than 100 lines of heterozygotic *AtFAP3*-RNAi were analysed; these consistently showed a restoration of the WT vegetative phenotype (Supplementary Fig. 22). However, the aberrant reproductive phenotype of the *AtFAP3*-RNAi mutant was maintained, indicating that phenotypic effects observed in the *AtFAP3*-RNAi lines are due to the specific knockdown of this gene. Analysis of fatty-acid composition reveals that *Atfap1* null lines have elevated total fatty-acid levels both in leaves and seeds relative to WT (Fig. 3d, e, g). These phenotypic alterations in leaves are temperature dependent (apparent only in plants grown at 15 °C and 22 °C) and are due primarily to increased levels of the lower-melting-temperature unsaturated FAs,

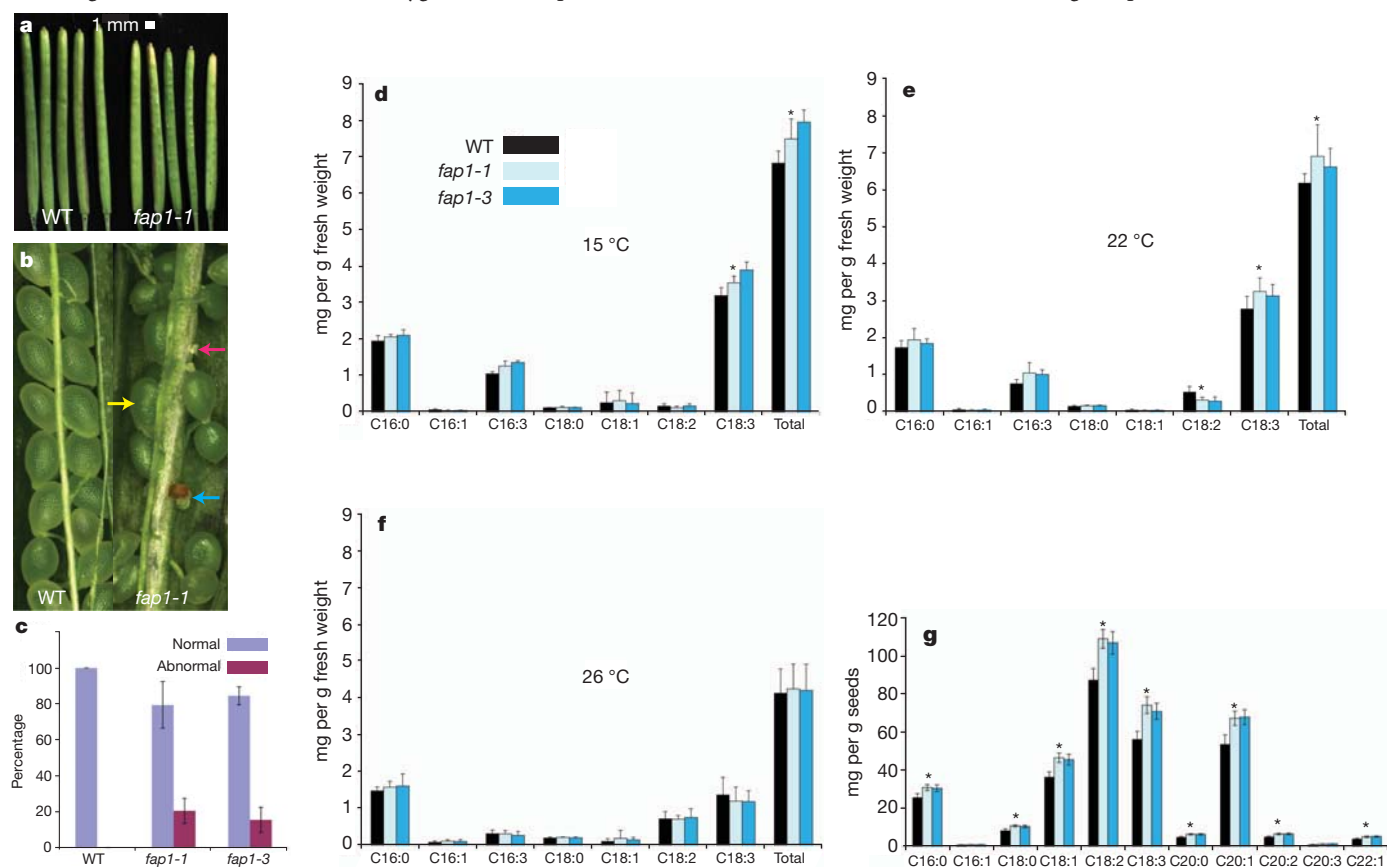


Figure 3 | Phenotypic characterization of *Atfap1* null plants. **a**, Length of siliques 12 days after flowering; WT siliques are longer than siliques of *Atfap1* nulls (average length = 17.8 mm for WT versus 15.4 mm for *Atfap1-1*; 19.8 mm for WT versus 17.1 mm for *Atfap1-3* (not shown); $P < 0.01$, $n = 10$ siliques per plant for 10 plants per genotype). **b**, *Atfap1* null siliques frequently contain abnormal ovules (normal, yellow arrow; aborted, blue arrow; unfertilized, pink arrow). **c**, *Atfap1* nulls have a greater percentage of abnormal (aborted and unfertilized) ovules than WT ($P < 0.01$ and $n = 10$ siliques per genotype) and low seed yield mass per plant (*Atfap1* average = 1.3 g; WT average = 1.5 g; $P < 0.01$,

$n = 10$ plants). Fatty-acid content of leaves, milligrams per gram fresh weight, from plants grown at **d**, 15 °C; **e**, 22 °C; **f**, 26 °C. **d**, C18:3 and total fatty acids are greater in leaves of *Atfap1* nulls than WT. **e**, C18:3 and total fatty acids are greater in leaves of *Atfap1* nulls than WT. **f**, *Atfap1* nulls and WT have similar fatty-acid content. **g**, In seeds of *Atfap1* mutants, C16:0, C18:0, C18:1, C18:2, C18:3, C20:0, C20:1, C20:2, C22:1 and total fatty acids (not shown) increase relative to WT. Asterisks indicate statistically significant differences from wild type for both *Atfap1* null lines ($P < 0.05$); $n = 5$ biological replicates, except for **g**, which has three experimental replicates with $n = 4$ biological replicates; error bars, s.d.

particularly C18:3. The fatty-acid content and distribution in leaves are indistinguishable between WT and mutant plants grown at 26 °C (Fig. 3f). Similarly, *AtFAP3*-RNAi plants and *Atfap* double mutants have altered fatty-acid composition and increased fatty-acid content in leaves and seeds (Supplementary Fig. 23). The temperature dependence of this phenotype reflects a decrease in the more fluid acyl chain component of trienoic fatty acids that accompanies increasing temperature^{16,17}.

We have identified three relatives of CHI in *Arabidopsis* that appear to be FAPs of undetermined mechanism. Phylogeny and sequence analyses indicate that the CHI fold occurs in most eukaryotic lineages (with the notable exception of animals) and even in bacteria. CHIL, as a divergent relative of FAP3, initially arose in mosses and may have served as the ancestor of enzymatic CHI in vascular plants (Supplementary Figs 2 and 3). This ancestry for plant CHI clarifies the mystery surrounding the identity of the progenitor protein, and serves as a model to understand adaptation of a ligand-binding protein to a highly efficient and stereospecific enzyme. An evolutionary connection between fatty acid and flavonoid metabolism¹⁸ is also evident from the structural and mechanistic similarities between chalcone synthase (a type-III polyketide synthase) acting one step upstream of CHI in the flavonoid pathway and ketoacyl-synthase III involved in chloroplast and bacterial fatty-acid biosynthesis. Together, CHI and CHS provide the foundation to examine the evolution of more recent metabolic pathways and metabolite classes from widespread biosynthetic pathways of primary metabolism.

METHODS SUMMARY

Structural biology. Crystal structures of the CHI domains of AtCHI, AtCHIL, AtFAP1 and AtFAP3 were determined as described in Methods and summarized in Supplementary Table 8.

In vitro biochemistry. CHI activity was assayed^{5,7}. Ligands bound to the purified AtFAP proteins were extracted with ethanol and analysed by reversed-phase HPLC-MS. The effects of fatty acids on the melting temperature of the AtFAP proteins were measured using a ThermoFluor-based assay¹².

Phylogenetic and sequence analyses. Homologues of AtCHI were identified by iterative PSI-BLAST and profile hidden Markov model. Alignments were curated using protein structure superpositions and homologous groups determined manually.

Informatics. Co-expression patterns of *A. thaliana* genes encoding FAPs and genes associated with fatty-acid biosynthesis were evaluated with data compiled from 72 microarray experiments using MetOmGraph (<http://www.metnetdb.org>)¹³.

Plant growth and mutant analyses. Plants were grown in randomized block designs at 22 °C under 16 h light/8 h dark. Fatty acids were measured 5 h after onset of light. Temperature effects were observed after 10 day growth at the specified temperatures. Homozygous *Atfap1-1*, *Atfap1-3*, *Atfap2-1* and *Atfap2-2* mutant alleles (Supplementary Fig. 14 and Supplementary Table 3) were generated from SALK_130560, SALK_039829, SAIL_171_C12 and SAIL_616_D09 (Arabidopsis Biological Resource Center) stocks, respectively. *Atfap3* RNAi lines (Supplementary Fig. 16 and Supplementary Table 9) were generated from *A. thaliana* Col-0 plants, transformed with *Agrobacterium tumefaciens* strain GV3101 containing vector CATMA1a44560 (Nottingham Arabidopsis Stock Center)¹⁹. *AtFAP* RNA was evaluated by quantitative RT-PCR (Supplementary Figs 15 and 16). Transgenic *A. thaliana* lines were generated with *AtFAP1*, -2, -3 promoters or promoter + cTP and/or + CDS fused to GUS and GFP.

Fatty-acid quantification in plants. Fatty acids were extracted from plant tissues using barium hydroxide hydrolysis²⁰, modified by methylation and analysed by gas chromatography-mass spectrometry (GC-MS). Nonadecanoic acid was used as an internal standard and a reference fatty-acid mixture was used for calibration and retention time determination (<http://www.plantmetabolomics.org>).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 April 2011; accepted 1 March 2012.

Published online 13 May 2012.

- Hartmann, T. From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry* **68**, 2831–2846 (2007).
- Ferrer, J.-L., Austin, M. B., Stewart, C. Jr & Noel, J. P. Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiol. Biochem.* **46**, 356–370 (2008).
- Peer, W. A. & Murphy, A. S. Flavonoids and auxin transport: modulators or regulators? *Trends Plant Sci.* **21**, 556–563 (2007).

- Bednar, R. A. & Hadcock, J. R. Purification and characterization of chalcone isomerase from soybeans. *J. Biol. Chem.* **263**, 9582–9588 (1988).
- Jez, J. M. & Noel, J. P. Reaction mechanism of chalcone isomerase: pH dependence, diffusion control, and product binding differences. *J. Biol. Chem.* **277**, 1361–1369 (2002).
- Albery, W. J. & Knowles, J. R. Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* **15**, 5631–5640 (1976).
- Jez, J. M., Bowman, M. E. & Noel, J. P. Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. *Biochemistry* **41**, 5168–5176 (2002).
- Jez, J. M., Bowman, M. E., Dixon, R. A. & Noel, J. P. Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nature Struct. Biol.* **7**, 786–791 (2000).
- Gensheimer, M. & Mushegian, A. Chalcone isomerase family and fold: no longer unique to plants. *Protein Sci.* **13**, 540–544 (2004).
- Shirley, B. W., Hanley, S. & Goodman, H. M. Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis* transparent testa mutations. *Plant Cell* **4**, 333–347 (1992).
- Oursel, D. *et al.* Identification and relative quantification of fatty acids in *Escherichia coli* membranes by gas chromatography/mass spectrometry. *Rapid Commun. Mass Spectrom.* **21**, 3229–3233 (2007).
- Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **9**, 2212–2221 (2007).
- Mentzen, W. I. & Wurtele, E. S. Regulon organization of *Arabidopsis*. *BMC Plant Biol.* **8**, 99 (2008).
- Mentzen, W. I., Peng, J., Ransom, N., Nikolau, B. J. & Wurtele, E. S. Articulation of three core metabolic processes in *Arabidopsis*: fatty acid biosynthesis, leucine catabolism and starch metabolism. *BMC Plant Biol.* **8**, 76–90 (2008).
- Ruuska, S. A., Girke, T., Benning, C. & Ohlrogge, J. B. Contrapuntal networks of gene expression during *Arabidopsis* seed filling. *Plant Cell* **14**, 1191–1206 (2002).
- Falcone, D. L., Ogas, J. P. & Somerville, C. R. Regulation of membrane fatty acid composition by temperature in mutants of *Arabidopsis* with alterations in membrane lipid composition. *BMC Plant Biol.* **4**, 17–31 (2004).
- Murakami, Y., Tsuyama, M., Kobayashi, Y., Kodama, H. & Iba, K. Trienoic fatty acids and plant tolerance of high temperature. *Science* **287**, 476–479 (2000).
- Austin, M. B. & Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **20**, 79–110 (2003).
- Hilson, P. *et al.* Versatile gene-specific tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications. *Genet. Res.* **14**, 2176–2189 (2004).
- Bonaventure, G., Salas, J. J., Pollard, M. R. & Ohlrogge, J. B. Disruption of the *FATB* gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. *Plant Cell* **15**, 1020–1033 (2003).
- DeLano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, 2002).
- Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Perera, B. Nikolau, H. Ilarsan and J. Peng for technical training (to M.N.N.), J. Peng for the *fap1-1* homozygote mutant line, D. Nettleton and H. Wang for statistical analysis of initial seed fatty-acid data, and Eric Scheeff for assistance in Bayesian phylogenetic analysis. This research was supported in part by a Fulbright Fellowship (to M.N.N.). This material is based in part upon work supported by the National Science Foundation under award number MCB-0645794 (to J.P.N.), EEC-0813570 (to E.S.W.), MCB-0951170 (to E.S.W.), and by National Cancer Institute award number CA14195 (to G.M.) and the Plant Sciences Institute at Iowa State University (to E.S.W.). J.P.N. is an investigator with the Howard Hughes Medical Institute. Portions of this research were conducted at the Advanced Light Source, a national user facility operated by Lawrence Berkeley National Laboratory, on behalf of the US Department of Energy, Office of Basic Energy Sciences. The Berkeley Center for Structural Biology is supported in part by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences. We thank the staff at the Advanced Light Source for assistance with X-ray data collection. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Author Contributions M.N.N. experimentally characterized the FAP genes in *planta*. M.E.B., R.N.P. and E.L. expressed, purified and crystallized proteins. L.L. designed genetics experiments and constructs. G.V.L. and F.P. performed fatty-acid binding analyses and solved the X-ray crystal structures. R.N.P. performed thermal-shift assays of fatty-acid binding. G.M. and E.L. performed phylogenetic and sequence analyses. J.P.N. designed the biochemical experiments; E.S.W. designed the bioinformatics and functional genomics experiments. The manuscript was written by R.N.P., G.V.L., M.N.N., J.P.N., G.M. and E.S.W.

Author Information Coordinates and structure factors are deposited in Protein Data Bank under accession numbers 4DOI (AtCHI), 4DOK (AtCHIL), 4DOL (AtFAP3) and 4DOO (AtFAP1). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.S.W. (mash@iastate.edu) or J.P.N. (noel@salk.edu).

METHODS

Phylogenetic and sequence analyses. Homologues of AtCHI were gathered by PSI-BLAST from the National Center for Biotechnology Information NR database, followed by iterative profile hidden Markov model building and searches against public protein, expressed sequence tag and genome sequence databases. Gene models were built with Genewise²³ and refined manually. Protein sequences in each class were aligned with MUSCLE²⁴ and manually edited, and classes were aligned by ClustalW²⁵ and further edited based on the structural superposition of the available structures. An alignment of all available protist and plant proteins and selected fungal and bacterial proteins is attached (Supplementary File 1). A consensus phylogenetic tree was built from 30,000 trees sampled from a 4 million-generation run of MrBayes²⁶ (Supplementary Fig. 2 and Supplementary File 2). A maximum likelihood tree build using PhyML²⁷ produced a highly similar branching pattern supporting the overall family classification, though several terminal branches were not conserved (not shown).

Expression and purification of *A. thaliana* CHI-fold proteins. The coding sequence for each *A. thaliana* protein's CHI-fold domain was inserted between the NcoI and BamHI sites of the expression vector pHis8, which, under the control of a T7 promoter, yields the target protein fused to a thrombin-cleavable amino-terminal His₈-tag²⁸. Proteins were heterologously expressed in *E. coli* BL21 (DE3) (Novagen) expression host. *E. coli* cultures were grown at 37 °C in Terrific Broth to an attenuation (600 nm) of 1.5, induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG), then grown for an additional 6 h at 22 °C. Bacterial cells were collected by centrifugation, re-suspended in lysis buffer (50 mM Tris-HCl, pH 8.0; 0.5 M NaCl; 20 mM imidazole; 1% Tween20; 10% glycerol; and 20 mM 2-mercaptoethanol) and lysed by sonication. Proteins were isolated from *E. coli* lysates by affinity chromatography with Ni-affinity (Ni²⁺-NTA) agarose, and eluted with lysis buffer containing 0.25 M imidazole. Partly purified proteins were treated with thrombin for cleavage of the His₈-tag, then further purified by gel-exclusion chromatography with a Superdex 200 HR26/60 column (Pharmacia Biosystems). Selenomethionine-substituted AtCHIL was generated from *E. coli* cultures grown in M9 minimal medium supplemented with an amino acid cocktail including L-selenomethionine²⁹ (Sigma-Aldrich), and were otherwise handled as described above for unlabelled protein.

Analysis of fatty-acid binding by *A. thaliana* CHI-fold proteins. Bound ligands were extracted from protein samples of AtFAP1, AtFAP2 and AtFAP3 (1–2 mg in 20 mM ammonium bicarbonate) by addition of ethanol to a final concentration of 80%. After incubation at –20 °C for 3 days, the treated samples were centrifuged (16,000g at 4 °C) to remove denatured protein and other particulates, and the supernatant was evaporated under vacuum. Residual material was re-dissolved in 200 µl 3-propanol, which was then passed through a 0.45 µm nylon filter in preparation for further characterization by HPLC–MS. The analysis of binding of exogenously provided α-linolenic acid by AtFAP1 used the aforementioned amino-terminal His₈-tagged form of the protein coupled to Ni-affinity resin. Immobilized protein (2 mg) was incubated in buffer (12.5 mM Tris-HCl, pH 8.0, 50 mM NaCl) with 0.5 mM α-linolenic acid, and subsequently washed thoroughly with buffer including 5% ethanol. The AtFAP1 protein was released with 0.25 M acetic acid, then processed for extraction of bound ligand as described above. Control samples were treated identically, except that the AtFAP1 protein was omitted.

Extracts were analysed for fatty-acid content by HPLC–MS on an Agilent 1100 Series LC–MSD instrument with electrospray-ionization eluent introduction into an XCT ion trap mass spectrometer (Agilent). Chromatographic separations employed an Agilent Zorbax Eclipse XDB–C18 (4.6 mm × 150 mm, 5 mm particle size) reversed-phase column run at a flow rate of 0.5 ml min^{–1}, and a linear gradient with initial and final mobile phases consisting of 95% water:5% acetonitrile:0.1% formic acid and 5% water:95% acetonitrile:0.1% formic acid, respectively. The identities of fatty-acid components were established by mass determination and comparison of chromatographic retention times with authentic fatty-acid standards (Sigma-Aldrich).

Thermal-shift binding assay. The shift in melting temperature for AtFAP proteins observed under increasing concentrations of fatty acids was measured using a Thermofluor-type assay similar to one previously published³⁰. Protein melting temperatures were assayed using the LightCycler480 System II (Roche), using the following program: 30 s at 20 °C, ramp up to 85 °C at 0.06 °C s^{–1}, 30 s at 20 °C, excitation wavelength 483 nm, emission wavelength 568 nm. Using SYPRO Orange (Sigma), an environmentally sensitive dye that interacts with hydrophobic amino-acid residues, the melting of a protein can be observed as a fluorescence increase when the hydrophobic core residues of a protein fold are exposed to the environment in a denatured protein. SYPRO Orange dye can interact with the unfolded protein and cause an increase in fluorescence. The maximum or minimum of a first derivative curve of the fluorescence profile will indicate the melting temperature (*T*_M). The melting temperature of a protein in the absence of any additional

compound provides a baseline melting temperature for the protein (*T*_M). Compounds can be screened for binding interactions with the protein of interest. When a ligand binds a protein, a change in Gibbs free energy occurs and may cause a change in observed *T*_M for that protein. Δ*T*_M can be measured for a collection of putative ligands and the strength of binding interactions can be ranked and compared for compounds with similar physicochemical properties, providing a relative measure of binding affinities.

A quantitative analysis of these relative binding affinities for AtFAPs versus the collection of fatty acids shown in Supplementary Table 1 (12 different fatty acids) can be obtained by a simple dose–response analysis of Δ*T*_M with respect to fatty-acid concentration. Five micromoles of each protein and 10× SYPRO Orange (from 5000× stock solution) were mixed with from 10^{–1} to 10⁵ nM of the 12 different fatty acids, incubated at room temperature for 20 min, then assayed for protein melting temperature. Mean Δ*T*_M (*T*_M – *T*_{M,measured}) was plotted against fatty-acid concentration for each FA/FAP combination using data from four replicate experiments. A 'dose–response: stimulation (three parameters)' standard (Hill) slope nonlinear regression was fit to the data, using this model: $Y = T_{M_{\min}} + [(T_{M_{\max}} - T_{M_{\min}})/(1 + 10^{(\log EC_{50} - X)})]$. Calculated values of EC₅₀ ± s.e.m. were reported as relative measures of binding constants, whereas maximum Δ*T*_M was reported as a qualitative measure of free-energy change in the protein + ligand system.

Crystallization and structure determination of *A. thaliana* CHI-fold proteins.

Crystals of AtCHI-fold proteins were grown by vapour diffusion at 4 °C, from 1:1 mixtures of protein solution (10–15 mg ml^{–1} in 12.5 mM Tris-HCl, pH 7.5, 50 mM NaCl) and reservoir solution. The reservoir solution contained 28% polyethylene glycol (PEG) 8000, 0.3 M magnesium nitrate, 2 mM dithiothreitol (DTT) and 100 mM HEPES-Na⁺ (pH 7.0) for AtCHI; 28% PEG 8000, 0.2 M calcium acetate, 2 mM DTT and 100 mM TAPS-Na⁺ (pH 8.5) for AtCHIL; 19% PEG 3350, 0.3 M potassium chloride, 2 mM DTT and 100 mM TAPS-Na⁺ (pH 8.5) for AtFAP1; and 7–9% PEG 8000, 0.2 M calcium acetate, 2 mM DTT and 100 mM PIPES-Na⁺ (pH 6.5) for AtFAP3. Crystal growth typically occurred over a period of 2–10 days, and was sometimes expedited through seeding with finely crushed microcrystals. For heavy-atom complexes, crystals of AtFAP1 were soaked overnight in mother liquor supplemented with 1 mM K₂PtCl₄.

Crystal samples were flash frozen by immersion in liquid nitrogen after a brief incubation in a cryoprotectant solution (consisting of reservoir solution supplemented with 17–20% ethylene glycol). X-ray diffraction data were collected from frozen crystals at the FIP beamline of the European Synchrotron Radiation Facility, beamlines 8.2.1 and 8.2.2 of the Advanced Light Source, Lawrence Berkeley National Laboratory, or beamlines 1-5 and 9-1 of the Stanford Synchrotron Radiation Laboratory. Diffraction intensities were measured on ADSC Quantum or MarResearch charge-coupled device detectors, and were indexed, integrated and scaled with MOSFLM³¹ and SCALA³², HKL2000³³, or XDS and XSCALE³⁴ programs.

Crystallographic structure solutions for the CHI-fold proteins were obtained through either single/multiple-wavelength anomalous dispersion or molecular-replacement analyses, as detailed in Supplementary Table 8. For single/multiple-wavelength anomalous dispersion analyses, the location of anomalous scatterers and initial phase estimates were determined with the program SOLVE³⁵, and preliminary structural models were automatically built with the program RESOLVE³⁶. For molecular-replacement analyses, Molrep³⁷ was used, and where necessary, search coordinate-sets were constructed through homology modelling with Modeller³⁸. ARP/wARP³⁹ was used for automated rebuilding of initial structure models for AtCHIL and AtFAP3. Subsequent structural refinements used CNS⁴⁰ or REFMAC⁴¹. Xfit⁴² and Coot⁴³ were used for graphical map inspection and manual rebuilding of atomic models. Programs from the CCP4 suite⁴⁴ were used for all other crystallographic calculations. Structural superpositions were calculated with SSM⁴⁵.

The identity of two ordered small-molecules bound in the active site of AtCHI was inferred to be nitrate because of the high concentration (0.3 M) of magnesium nitrate in the crystallization medium. From the X-ray crystallographic analysis, the shape and level of the electron density associated with these small molecules are also consistent with nitrate.

Plant growth. This study used WT *A. thaliana* Columbia (Col-0 and Ws), and mutant and transgenic lines derived from them. Seeds (approximately five per pot) were sown on soil, in flats containing either 21 or 32 pots; genotypes were distributed in a completely randomized design. After incubation for 3 days at 4 °C to break dormancy, flats were moved to a growth-chamber under long-day conditions (16:8 hours light:dark cycle) at 22 °C and 75% relative humidity; after a week pots were thinned such that each pot contained two plants. For additional plant morphological analyses and for seed collection, plants were grown in the greenhouse at 20–23 °C under continuous illumination at (170 µmol m^{–2} s^{–1}) in soil

treated with granular Marathon (Olympic Horticultural Products) to protect against insect damage.

In the specified experiments, after 3 weeks of growth at 22 °C, plants were further grown at 15, 22 and 26 °C until senescence (control plants were grown at 22 °C for this period). Genotypes were distributed in a randomized design. Fatty-acid determinations were made from expanded rosette leaves harvested after 10 days of growth at 15, 22 and 26 °C. Growth and morphological phenotype of plant lines were observed every 2 days throughout development, and any differences between WT and mutants were recorded.

Molecular constructs, transgenic *A. thaliana* lines and microscopy. To evaluate subcellular location of the proteins, we generated promoter + target-GUS/GFP fusion constructs using promoters of two different lengths for *FAP1* and *FAP3*, and promoter + CDS-GUS/GFP fusion for *FAP2*. The primers used for amplification were as follows: *FAP1* prom1 + T FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTCTTCACTTCAAC-3'); *FAP1* prom2 + T FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTCTTCACTTCAAC-3'); *FAP1* prom + T Rev 5'-GGGGACCACCTTTGTACAAAGAAAGCTGGGTTACGATCACACGGGAAATTCG-3'); *FAP2* prom + CDS FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTCTTCACTTCAAC-3'); *FAP2* prom + CDS Rev (5'-GGGGACCACCTTTGTACAAAGAAAGCTGGGTTTATAAGCTTTGAATATAACTCCA-3'); *FAP3* prom1 + T FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTCTTCACTTGAACCAATAACTAGCAA-3'); *FAP3* prom2 + T FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTCTTCACTTGAACCAATAACTAGCAA-3'); *FAP3* prom + T Rev (5'-GGGGACCACCTTTGTACAAAGAAAGCTGGGTTGTACAAACTCCGGCGATAA-3'). These constructs harboured GFP fused to the amino-terminal region of *FAP1/3* genes and the CDS of *FAP2*, under the control of the specified promoter. The constructs were cloned into the pDONR221 entry vector, integrated into the binary vector pBGWFS7 (ref. 46), and transformed into *A. tumefaciens* strain GV3101 by electroporation, which we used to transform *A. thaliana* (Col-0) by the 'floral-dip' method⁴⁷. After identification of putative transgenic *T₁* lines by bar resistance⁴⁸, plants from 10 independent lines for each construct were analysed by PCR using the attB-adapted primers (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTC-forward primer, 5'-GGGGACCACCTTTGTACAAAGAAAGCTGGGTC-reverse primer). Five independent transgenic lines in the *T₂* generation (five individual plants from each transgenic line at each stage of development) were evaluated for GFP location by confocal microscopy⁴⁹.

To evaluate the spatial distribution of FAP expression *in planta*, we generated promoter-GUS/GFP fusion constructs using promoters of two different lengths. The primers used for amplification were as follows: *FAP1* prom1 FW (5'-GGGACAAGTTTGTACAAAAAGCAGGCTACTGGTAACTTCAATTAACTTCCT-3'); *FAP1* prom2 FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTGTCTTCACTGAAATTATCTTCAAC-3'); *FAP1* prom Rev (5'-GGGGACCACCTTTGTACAAAGAAAGCTGGGTCGTCGCTTTTGGTGTGAG-3'); *FAP2* prom FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTAGAAAGTGAACACACACAAAA-3'); *FAP2* prom Rev (5'-GGGGACCACCTTTGTACAAAGAAAGCTGGGTTTATAAGCTTTGAATATAACTCCA-3'); *FAP3* prom1 FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTCTTCACTTGAACCAATAACTAGCAA-3'); *FAP3* prom2 FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTCTTCACTTGAACCAATAACTAGCAA-3'); *FAP3* prom Rev (5'-GGGGACCACCTTTGTACAAAGAAAGCTGGGTTTATAAGCTTTGAATATAACTCCA-3'). Cloning, transformation and selection methods were as previously described in this section. Ten independent lines were obtained for each construct and analysed by PCR. Representative transgenic lines in the *T₂* generation were subjected to histological tissue analysis.

In an additional set of constructs for determining subcellular location, the full open reading frame (ORF) of *FAP1* and *3* was introduced into pEarleyGate 103 (ref. 50), to generate an in frame fusion with GFP under the control of the cauliflower mosaic virus 35S promoter. The cloning procedure was similar to the one described above. The primers used for amplification of the open reading frames were *FAP1*p103 FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTCATGGTTTCGTTTCGCTTCC-3'); *FAP1*p103 Rev (5'-GGGGACCACCTTTGTACAAAGAAAGCTGGGTTGTGTGCAAGGCTAGAGAAGC-3'); *FAP3*p103 FW (5'-GGGGACAAGTTTGTACAAAAAGCAGGCTTCTTCACTGATGGGATGCTTCTGCAG-3'); *FAP3*p103 Rev (5'-GGGGACCACCTTTGTACAAAGAAAGCTGGGTTCACTAGGGTCATGGCTAATTG-3'). After insertion of FAPs open reading frame in pDONR221, which contains kanamycin resistance (bacterial selection marker also found in the binary vector pEarleygate 103), we cleaved the transformed pDONR221 using restriction enzymes that do not cut the inserted sequence and purified the fragment of interest. Transgenic plants bearing promoter + target-GFP (cTP-GFP) gave indistinguishable results from those bearing 35S + CDS-GFP.

Near fully expanded leaves from 2-week-old *T₂* mutant plants (five independent lines) were collected for confocal microscopy observation using a Leica TCS NT laser scanning microscope system (Confocal Microscopy Facility, Iowa State University). For each line, five plants were analysed. GFP signal and auto fluorescence were detected at wavelengths of 488 nm FITC and 568 nm TRITC, respectively (Supplementary Fig. 4). We imported the images into Adobe Photoshop version 6.0 (Adobe).

For histochemical studies, transgenic and control plants (siblings not containing the promoter-GUS constructs) were collected at different stages of development. Plant material was incubated at room temperature in the dark with a GUS⁵¹ substrate solution containing 10 mg ml⁻¹ X-Gluc in dimethylsulphoxide, 0.1 M potassium phosphate buffer (pH 7.0), Triton/ethanol (Triton X-100:ethanol:water; 1:4:5), 0.1 M K₃[Fe(CN)₆] (pH 7.0), 0.1 M K₄[Fe(CN)₆] (pH 7.0), washed with 70% ethanol, observed and photographed with an Olympus stereomicroscope⁴⁹ (Bessey Microscopy Facility, Iowa State University).

T-DNA insertion lines and mutant isolation. Seeds of *AtFAP1-1* (*Salk_130560*) and *AtFAP1-3* (*Salk_039829*) were obtained from the Arabidopsis Biological Resource Center; *AtFAP1-2* (*Flag_389_G05*) seeds were obtained from the Institut National de la Recherche Agronomique. PCR reactions with a combination of specific primers LP + RP and LB + RP were used to verify the site of T-DNA insertion for each line (Supplementary Table 9). The T-DNA insertion in *AtFAP1-1* is located in the first intron; in *AtFAP1-2* and *AtFAP1-3*, it is in the second exon (Supplementary Fig. 14). Knockout mutants were identified as homozygous via PCR and by backcrossing to WT and mutant parents (once each; heterozygous crosses resulted in 1:1 segregation pattern).

We previously determined that the putative *AtFAP3* mutant lines available from SALK and SAIL do not have the T-DNA insertion in *AtFAP3*. Specifically, we came to this conclusion after ordering and testing multiple batches of seeds of the SALK_107742 and SAIL_598_D12 putative *AtFAP3* mutant lines. The plants from these lines were screened by PCR using the primer combinations as suggested in TAIR ((SALK_107742-LP + SALK_107742-RP, lbba1/lbba1-3 + SALK107742-RP); (SAIL_598-LP + SAIL_598-RP, LB2/LB3 + SAIL_598-RP)), in parallel with controls (SALK/SAIL lines representing the other FAP mutants). We were unable to identify either homozygote mutant or heterozygote mutant plants in the SALK_107742 or SAIL_598 lines. Unlike the multiple independent *AtFAP3-RNAi* lines we generated, the SALK_107742 or SAIL-598 lines were visibly indistinguishable from those of the corresponding WT.

DNA and RNA isolation from plant tissue and RT-PCR. Genomic DNA was isolated from leaves using a cetyl trimethylammonium bromide extraction protocol⁵². Tissue was frozen and ground to a powder in liquid nitrogen. Buffer composed of 1% cetyl trimethylammonium bromide, 50 mM Tris-HCl, pH 8.0, 0.7 mM NaCl, 10 mM EDTA, 0.5% polyvinylpyrrolidone (PVP), was added; samples were homogenized and incubated at 65 °C for 1 h. The solution was extracted with chloroform, followed by isopropanol precipitation of DNA. After incubation for 10–30 min at –20 °C, the pellet was treated with RNase before precipitation with sodium acetate, centrifugation and washing with 70% ethanol. After centrifugation, the pellet was dried, re-suspended in water, and DNA was amplified by PCR as previously described.

Total RNA was isolated from leaves from 2-week-old WT and mutant plants with TRIzol reagent (Invitrogen) following the manufacturer's instructions. Isolated RNA was treated with amplification-grade DNaseI (Invitrogen) to remove genomic DNA in the samples. The purified RNA was used for RT-PCR amplification with the SuperScript III One-Step Kit (Invitrogen) according to the manufacturer's instructions. The forward and reverse primers used were as follows: 5'-ATGGTTTCGTTTCGCTTCC-3' and 5'-CTAGTTGTGCAAGGCTAGAGAAGCT-3', respectively; tubulin primers (FW 5'-CGTGGATCACAGCAATACAGAGCC-3' and REV 5'-CCTCCTGCATCTCCACTTCGTTCTTC-3') were used as a control for RNA content. The thermal profile used was as follows: one cycle: 45–60 °C for 15–30 min (complementary DNA synthesis); one cycle: 94 °C for 2 min (denaturation); 28 cycles: 94 °C for 15 s (denature), 56 °C for 30 s (anneal), 68 °C for 1 min kb⁻¹ (extend). The final extension comprised one cycle: 68 °C for 5 min.

Western blot analysis. Total protein was extracted from 3-week-old plant leaves and seven or eight DAF siliques; protein concentration was determined using the method of Bradford⁵³. Extracts containing 100 µg of total protein were loaded in each well for SDS-PAGE; for a positive control, 0.1 µg of purified protein was used. Proteins were transferred to nitrocellulose membrane using Bio-Rad apparatus. FAP3-specific antiserum was generated in rabbits at the Protein Facility of Iowa State University by using FAP3 protein expressed and purified from *E. coli*.

Fatty-acid analysis from plant tissue. Fatty acids were extracted using a barium hydroxide hydrolysis protocol^{20,54,55}. Recently expanded rosette leaves were collected and placed in liquid nitrogen. Each replicate comprised leaves from a single plant. Approximately 0.1 g fresh weight of leaves or 5 mg of seeds was used for each

extraction. Tissues were frozen in liquid nitrogen upon collection, and placed in a pre-cooled tissue homogenizer. Twenty microlitres of internal standard (nonadecanoic acid, C19:0, 2 mg ml⁻¹ dissolved in chloroform) and 1 ml of barium hydroxide were added, and the mixture was further homogenized according to the procedure in plant metabolomics.org (<http://www.plantmetabolomics.org>). The homogenate was transferred to a glass tube and 550 µl of 1,4-dioxane (Aldrich catalogue number 123-91-1) added. The tube was tightly capped and incubated for 24 h at 110 °C. The solution was then acidified with six drops of 6 M HCl, hexane (2 × 3 ml) was added, and the mixture was centrifuged. The hexane layer was transferred to a new tube, dried under nitrogen gas, methylated with 2 ml of HCl: methanol for 1 h at 80 °C, followed by two extractions in hexane (2 ml each). Samples were dried under nitrogen gas, acetylated with 1 ml of acetonitrile and 70 µl of bis-(trimethylsilyl)trifluoroacetamide for 20 min at 60 °C. The solvent was evaporated under nitrogen gas and 200 µl (for leaf samples) or 1.5 ml (for seed samples) of chloroform was added.

Chloroform-dissolved material was subjected to fatty-acid methyl ester analysis using a Model 6890 series gas chromatograph (Agilent) equipped with a Mass Detector 5973 (Agilent) and an HP-1 silica capillary column (30 m × 30.32 mm, inner diameter), using helium as a carrier gas^{56,57}. The injector and detector temperature was 300 °C. The initial oven temperature was 100 °C for 2 min, increased to 240 °C at a rate of 5 °C min⁻¹, and maintained at 240 °C for 5 min. The total running time was 40 min. AMDIS software version 2.65 was used for peak integration⁵⁸; peaks were identified through the Agilent NIST05 mass spectra libraries (<http://www.nist.gov/srd/nist1a.htm>). Fatty-acid methyl esters were quantified in each sample in accordance with the amount of nonadecanoic acid internal standard added.

Statistical and bioinformatics analysis. All experiments were conducted a minimum of three times. For each experiment, plants were collected and analysed in randomized block design. Data are presented as mean ± s.d. We compared two sets of independent samples using Student's *t*-test (two-tailed) with assumption of equal variances and *P* < 0.05 was considered significant.

Pearson correlation coefficients of *A. thaliana* transcript accumulation across changes in plant genotype, environment and development were calculated using MetaOmGraph (<http://www.metnetdb.org>)¹³ and a data set consisting of 951 chips from 72 public microarray experiments^{14,49}.

23. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
24. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
25. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
26. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755 (2001).
27. Guindon, S., Lethiec, F., Duroux, P. & Gascuel, O. PHYML Online – a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* **33**, W557–W559 (2005).
28. Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A. & Noel, J. P. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **39**, 890–902 (2000).
29. Doublié, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
30. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **9**, 2212–2221 (2007).
31. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* No. 26 (1992).
32. Evans, P. R. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2005).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
34. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800 (1993).
35. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
36. Terwilliger, T. C. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. D* **59**, 38–44 (2003).
37. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
38. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
39. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
40. Brunger, A. T. & Warren, G. L. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
41. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
42. McRee, D. E. XtalView/Xfit: a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
43. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
44. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
45. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60**, 2256–2268 (2004).
46. Invitrogen. Gateway technology: a universal technology to clone DNA sequences for functional analysis and expression in multiple systems. Catalog nos. 12535-019 and 12535-027. <http://www.wolfson.huji.ac.il/expression/gatewayman.pdf> (2003).
47. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
48. Lohar, D. P., Shuller, K., Buzas, D. M., Gresshoff, P. M. & Stiller, J. Transformation of *Lotus japonicum* using the herbicide resistance *bar* gene as a selectable marker. *J. Exp. Bot.* **52**, 1697–1702 (2001).
49. Li, L., Ilarslan, H., James, M. G., Myers, A. M. & Wurtele, E. S. Genome wide co-expression among the starch debranching enzyme genes *AtISA1*, *AtISA2*, and *AtISA3* in *Arabidopsis thaliana*. *J. Exp. Bot.* **58**, 3323–3342 (2007).
50. Earley, K. W. *et al.* Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).
51. Jefferson, R. A. Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol. Biol. Rep.* **5**, 387–405 (1987).
52. Doyle, J. J. & Doyle, J. L. Isolation of plant DNA from fresh tissue. *Focus* **12**, 13–15 (1990).
53. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Annu. Rev. Biochem.* **72**, 248–254 (1976).
54. Markham, J. E., Li, J., Cahoon, E. B. & Jaworski, J. G. Separation and identification of major plant sphingolipid classes from leaves. *J. Biol. Chem.* **281**, 22684–22694 (2006).
55. Moon, S. & Nikolau, B. J. Plantmetabolomics Platform2: fatty acids. <http://www.plantmetabolomics.org> (2009).
56. Perera, M. A. D. N., Dietrich, C. R., Meeley, R., Schnable, P. S. & Nikolau, B. J. in *Advanced Research on Plant Lipids* (eds Murata, M. *et al.*) 225–228 (Kluwer Academic, 2003).
57. Fatland, B. L., Nikolau, B. J. & Wurtele, E. S. Reverse genetic characterization of cytosolic acetyl-CoA generation by ATP-citrate lyase in *Arabidopsis*. *Plant Cell* **17**, 182–203 (2005).
58. Davies, N. A. *The New Automated Mass Spectrometry Deconvolution and Identification System (AMDIS)* (ISAS, 1998).

The complex of tmRNA–SmpB and EF–G on translocating ribosomes

David J. F. Ramrath^{1*}, Hiroshi Yamamoto^{1,2*}, Kristian Rother³, Daniela Wittek², Markus Pech^{1,2}, Thorsten Mielke^{1,4}, Justus Loerke¹, Patrick Scheerer¹, Pavel Ivanov^{2,5†}, Yoshika Teraoka², Olga Shpanchenko^{2,5}, Knud H. Nierhaus^{1,2} & Christian M. T. Spahn¹

Bacterial ribosomes stalled at the 3' end of malfunctioning messenger RNAs can be rescued by transfer-messenger RNA (tmRNA)-mediated *trans*-translation^{1,2}. The SmpB protein forms a complex with the tmRNA, and the transfer-RNA-like domain (TLD) of the tmRNA then enters the A site of the ribosome. Subsequently, the TLD–SmpB module is translocated to the P site, a process that is facilitated by the elongation factor EF–G, and translation is switched to the mRNA-like domain (MLD) of the tmRNA. Accurate loading of the MLD into the mRNA path is an unusual initiation mechanism. Despite various snapshots of different ribosome–tmRNA complexes at low to intermediate resolution^{3–7}, it is unclear how the large, highly structured tmRNA is translocated and how the MLD is loaded. Here we present a cryo-electron microscopy reconstruction of a fusidic-acid-stalled ribosomal 70S–tmRNA–SmpB–EF–G complex (carrying both of the large ligands, that is, EF–G and tmRNA) at 8.3 Å resolution. This post-translocational intermediate (TI^{POST}) presents the TLD–SmpB module in an intrasubunit ap/P hybrid site and a tRNA^{fMet} in an intrasubunit pe/E hybrid site. Conformational changes in the ribosome and tmRNA occur in the intersubunit space and on the solvent side. The key underlying event is a unique extra-large swivel movement of the 30S head, which is crucial for both tmRNA–SmpB translocation and MLD loading, thereby coupling translocation to MLD loading. This mechanism exemplifies the versatile, dynamic nature of the ribosome, and it shows that the conformational modes of the ribosome that normally drive canonical translation can also be used in a modified form to facilitate more complex tasks in specialized non-canonical pathways.

The tmRNA–SmpB-mediated *trans*-translation is a bacterial emergency system that rescues ribosomes stalled at the 3' end of defective mRNAs that lack a stop codon. The tmRNA is a highly structured bifunctional molecule that acts as tRNA and mRNA through its TLD and MLD^{1,2} (Fig. 1a). SmpB interacts with the TLD and mimics the tRNA anticodon stem that is lacking in the TLD^{8–10}. After the TLD–SmpB module has entered the A site of the ribosome in a decoding-like reaction, the immature polypeptide is transferred to Ala–tmRNA (tmRNA with the amino acid alanine loaded at its 3' end). Subsequently, the TLD–SmpB module is translocated to the P site of the ribosome, and the template for translation is switched to the MLD. The MLD encodes a tag, which marks the immature peptide for protease-dependent degradation and ends with a stop codon allowing canonical termination^{1,2} (Fig. 1a).

Several cryo-electron microscopy (cryo-EM) maps of ribosomal complexes with bound tmRNA have provided important information about the passage of tmRNA through the ribosome. At low to intermediate resolution, these maps display the TLD–SmpB module in the A/T site of the ribosome when the tmRNA is delivered to the ribosome as an Ala–tmRNA–SmpB–EF–Tu–GTP complex^{3,4}. The maps also

show the complexes after accommodation^{5,6} and first translocation^{6,7}, with the TLD–SmpB module occupying the ribosomal A and P sites, respectively. However, as discussed in previous reports^{11,12}, for a detailed interpretation of cryo-EM maps in molecular terms, subnanometre resolution is required so that secondary structural elements can be directly observed. Moreover, the interplay between the ribosome and its ligands during translocation, and the process by which the large and highly structured tmRNA passes through the ribosome and leads to correct MLD loading, remains unclear.

Here we present a cryo-EM reconstruction of an *in vitro*-reconstituted ribosomal 70S–tmRNA–EF–G complex from *Escherichia coli* (Fig. 1). Pre-translocated (PRE) ribosomes carrying fMet–Ala–tmRNA–SmpB in the A site and initiator tRNA^{fMet} in the P site were translocated by EF–G, which was stalled on the complex using the antibiotic fusidic acid. Because P-site-bound initiator tRNA^{fMet} has a slower rate of translocation than P-site-bound elongator tRNA^{13,14}, the usage of a tRNA^{fMet} might increase the time window for fusidic acid to trap the reaction. The resultant complex was analysed by multiparticle cryo-EM¹⁵ to overcome sample heterogeneity caused by substoichiometric binding of the ligands. The final map, reconstructed from a subpopulation of 68,842 particle images (9.5% of the total data set), reached the subnanometre resolution of 8.3 Å (Supplementary Fig. 1). The structure shows well-defined densities for the ligands fMet–Ala–tmRNA–SmpB, deacylated tRNA^{fMet} and EF–G. In agreement with the subnanometre resolution estimate, the structure contains clearly visible structured RNA or rod-shaped α -helical elements, and it readily allows the docking of atomic models of the ribosome¹⁶, EF–G–GDP–fusidic acid¹⁷, tRNA^{fMet} (ref. 18) and the TLD–SmpB module⁹ (Figs 1c and 2a). Moreover, we attempted RNA modelling of major tmRNA regions for which no high-resolution structure was available (Supplementary Table 1).

Compared with the ribosome in the classical state¹⁶, the 70S–tmRNA–EF–G complex has undergone large-scale conformational changes. The body/platform domains of the 30S subunit are rotated by $\sim 5^\circ$, and the head is swivelled by an additional $\sim 19^\circ$ (Fig. 2b). Thus, the conformation of the presented 70S–tmRNA–EF–G complex is similar overall to the recently described translocation intermediate TI^{POST}, which is characterized by an intermediate ($\sim 4^\circ$) 30S rotation and a large ($\sim 18^\circ$) head swivel¹⁹. However, the overall movement of the head is unique: compared with the head position of the previously reported TI^{POST} (ref. 19), there is an additional $\sim 12^\circ$ tilt movement around an axis that is approximately parallel to the mRNA, around the A- and P-site codons (Fig. 2b). This tilt moves the head away from the central protuberance of the 50S subunit. The combination of intersubunit rotation, head swivel and head tilt lead to a displacement of peripheral elements of up to ~ 50 Å compared with the classical state¹⁶ (Fig. 2b).

The two tRNA/tRNA-like entities are trapped in intermediate states during translocation. From the perspective of the 50S subunit and the

¹Institut für Medizinische Physik und Biophysik, Charité – Universitätsmedizin Berlin, Ziegelstrasse 5-9, 10117 Berlin, Germany. ²Max Planck Institute for Molecular Genetics, Abteilung Vingron, AG Ribosomen, Ihnestrasse 73, 14195 Berlin, Germany. ³Institute of Biology and Molecular Biotechnology, Collegium Biologicum, Adam Mickiewicz University, Ulica Umultowska 89, 61-614 Poznan, Poland. ⁴UltraStrukturNetzwerk, Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany. ⁵Department of Chemistry, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia. [†]Present address: Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

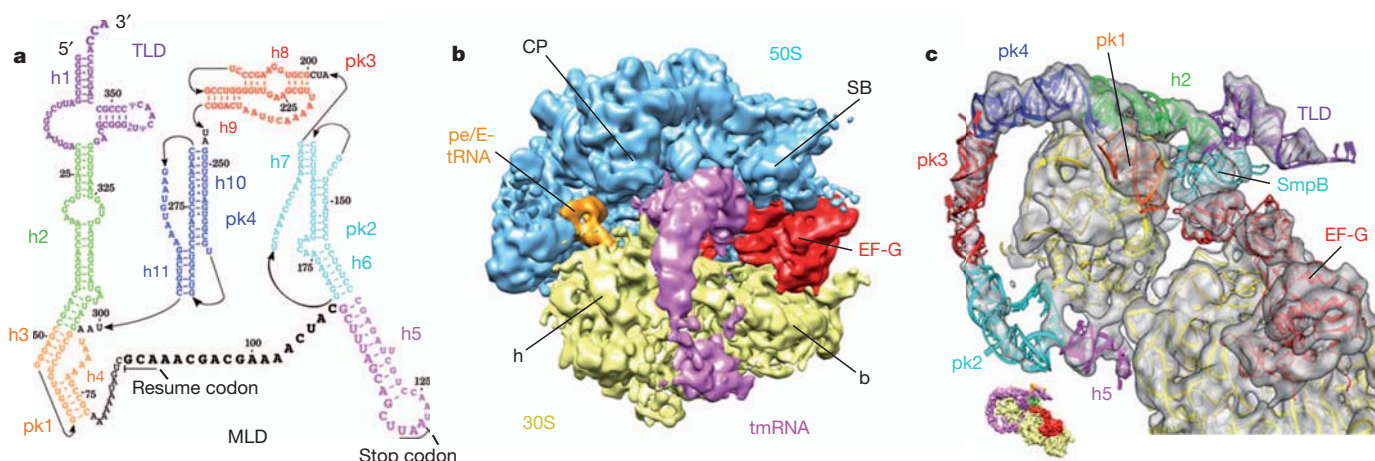


Figure 1 | Overall view of the 70S–tmRNA–EF-G complex. **a**, Secondary structure of the *E. coli* tmRNA with its TLD and MLD, helices h1 to h11, and the pseudoknots pk1 to pk4. **b**, Overall view of the cryo-EM reconstruction of the 70S–tmRNA–EF-G complex displaying the 50S subunit (blue), 30S subunit (yellow), EF-G (red), tmRNA (pink) and pe/E-tRNA (orange). **c**, View onto the 30S subunit with bound ligands as indicated by the orientation aid (bottom left), shown as a cartoon representation (coloured) of EF-G and a pseudo-atomic model of the ribosome-bound tmRNA–SmpB complex modelled onto the 8.3 Å resolution cryo-EM map (grey).

30S body, the fMet-Ala-TLD–SmpB module and the deacylated tRNA^{fMet} are bound to the P and E sites, respectively. However, the fMet-Ala-TLD–SmpB module still maintains A-site contacts with the swivelled 30S head, and the deacylated tRNA^{fMet} still maintains P-site contacts with the swivelled 30S head (Fig. 2a and Supplementary Fig. 2). Accordingly, they are bound in intrasubunit ap/P and pe/E hybrid sites (the lowercase and uppercase letters indicate tRNA contacts on the small and large subunit, respectively, in the following order: 30S head (a or p), 30S body (p or e) and 50S subunit (P or E)). It should be noted that our previous TI^{POST} contained only one tRNA, which was present in a pe/E hybrid state¹⁹. Thus, because fMet-Ala-TLD–SmpB is observed here in an ap/P state, this corroborates our recent model for canonical translocation¹⁹ and suggests that tmRNA and tRNA are translocated by the same mechanism (Supplementary Information and Supplementary Fig. 3). The antibiotic fusidic acid seems to have trapped the 70S–tmRNA–EF-G complex in an intermediate state during the translocation reaction, with the TLD bound to the 50S P site, in agreement with a toeprint signal from a fusidic-acid-trapped post-translocated (POST) state²⁰.

In canonical POST complexes, domain IV of EF-G comes into contact with the anticodon loop of the P-tRNA and intercalates into the minor groove of the anticodon–codon RNA helix¹⁷. In the present TI^{POST} , the overall position and conformation of EF-G is strikingly similar, and its domain IV comes into contact with SmpB around the 30S decoding centre (Fig. 3a, b). Thus, SmpB replaces the binding

interface of the canonical EF-G interaction partner during translocation, by mimicking the tRNA anticodon stem–loop. Likewise, several polar or aromatic SmpB residues around the binding interface are conserved across bacteria⁹, indicating the importance of this region for SmpB function (Fig. 3a and Supplementary Table 2).

As shown by previous cryo-EM reconstructions of fully translocated tmRNA^{6,7}, tmRNA maintains an ‘arc and fork’ structure, which includes pseudoknot 1 (pk1) and connects to the TLD by way of helix 2 (tmRNA-h2) (Figs 1c and 4a). Our structure reveals how conformational changes within the ribosome open up the intersubunit bridge B1a to allow the passage of folded tmRNA. A huge gap of 50 Å is formed between the B1a-forming elements S13 and the A-site finger (ASF) in the small and large subunit, respectively (Fig. 4a–c). The key to opening the bridge seems to be the extra-large 30S head movement that includes the unique tilt. Formation of the gap is aided by a lateral displacement of the ASF by up to 20 Å (Fig. 4c) compared with ribosomes with classically bound tRNAs²¹. This displacement shifts the ASF to a position where it is stabilized by a TI^{POST} -specific interaction with tmRNA pk1 (Fig. 4a and Supplementary Fig. 4).

On the solvent side of the 30S subunit, the opposite base of the tmRNA (in tmRNA-h5 and pk2) interacts with the ribosomal proteins S2 and S3 (Fig. 4a and Supplementary Table 2). Like the ASF–pk1 contact, the tmRNA-h5–S2 interaction has not been observed in other 70S–tmRNA complexes^{3–7} and thus seems to be another specific feature of the TI^{POST} . The tmRNA region connecting the 3′ MLD, tmRNA-h5

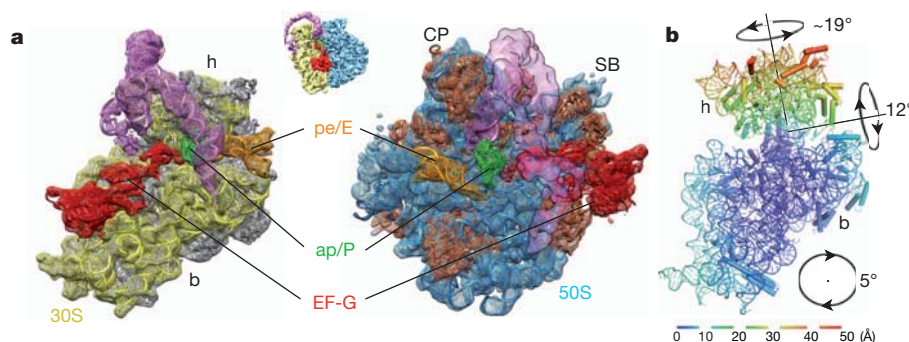


Figure 2 | Ligand positions and conformation of the 30S subunit. **a**, Mesh representation of the cryo-EM maps and docked models showing the ribosomal ligands (EF-G (red), tmRNA–SmpB (pink/green) and tRNA (orange)) relative to the 30S subunit (ribosomal RNA in yellow and S proteins in grey) (left) and the 50S subunit (rRNA in blue and L proteins in orange) (right) within the

TI^{POST} . **b**, Model of the 30S subunit in ribbon representation. Elements are coloured according to their structural displacement compared with the classical conformation¹⁶. The directions and magnitudes of the intersubunit rotation (5°), head swivel (19°) and tilt (12°) movements are indicated.

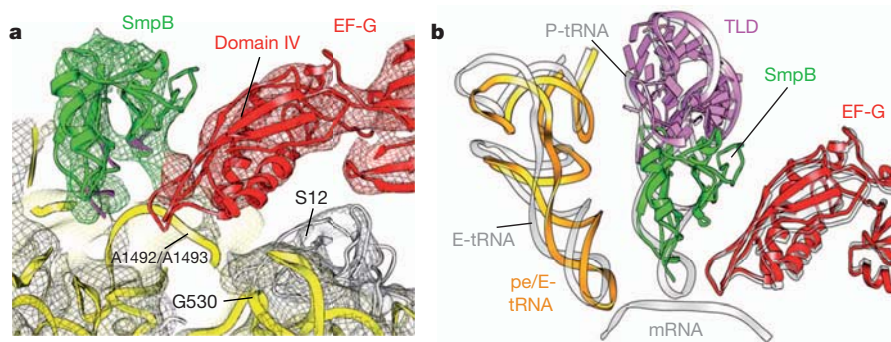


Figure 3 | The EF-G-SmpB interaction. **a**, Mesh representation of cryo-EM maps with docked models showing the interaction between EF-G (red) and SmpB (green) around the 30S decoding centre formed by A1492 (adenosine at position 1492), A1493 and G530 of 16S rRNA (30S coloured as in Fig. 2a). The positions of several conserved SmpB residues around the binding interface are indicated in pink⁹. **b**, Superimposition of the ribosomal ligands in the $T1^{POST}$ (EF-G (red), ap/P-TLD-SmpB (pink/green) and pe/E-tRNA (orange)), compared with ligand positions in a classical POST state¹⁷ (transparent grey), on a 50S alignment.

and pk2 is held in place by a highly conserved helix–turn–helix motif in S3 (Supplementary Fig. 5) and acts as a pivot point for a conformational change, bending tmRNA-h5 to an almost perpendicular orientation relative to pk2 (Fig. 4a). Comparison of the tmRNA positions in the $T1^{POST}$ and the accommodated state⁶ shows that, in addition to the TLD movement within the intersubunit space during translocation, there is a pronounced displacement of tmRNA elements (tmRNA-h5 and pk2), which interact with the solvent side of the 30S subunit (Fig. 4f). Thus, the tmRNA molecule is moved inside and outside the ribosome in

discrete steps. The interactions of tmRNA-h5 and pk2 with S3 are established in the preceding accommodated state⁶. This suggests that both elements follow S3 during the movement of the head, which explains how the conformational changes of the ribosome during translocation are linked to tmRNA movement on the solvent side.

During the switch from the malfunctioning mRNA to the MLD, the tmRNA has to accomplish an initiation-like task by placing the resume codon directly at the 30S decoding centre and threading the MLD into the ribosomal mRNA entry channel. One dynamic region of the

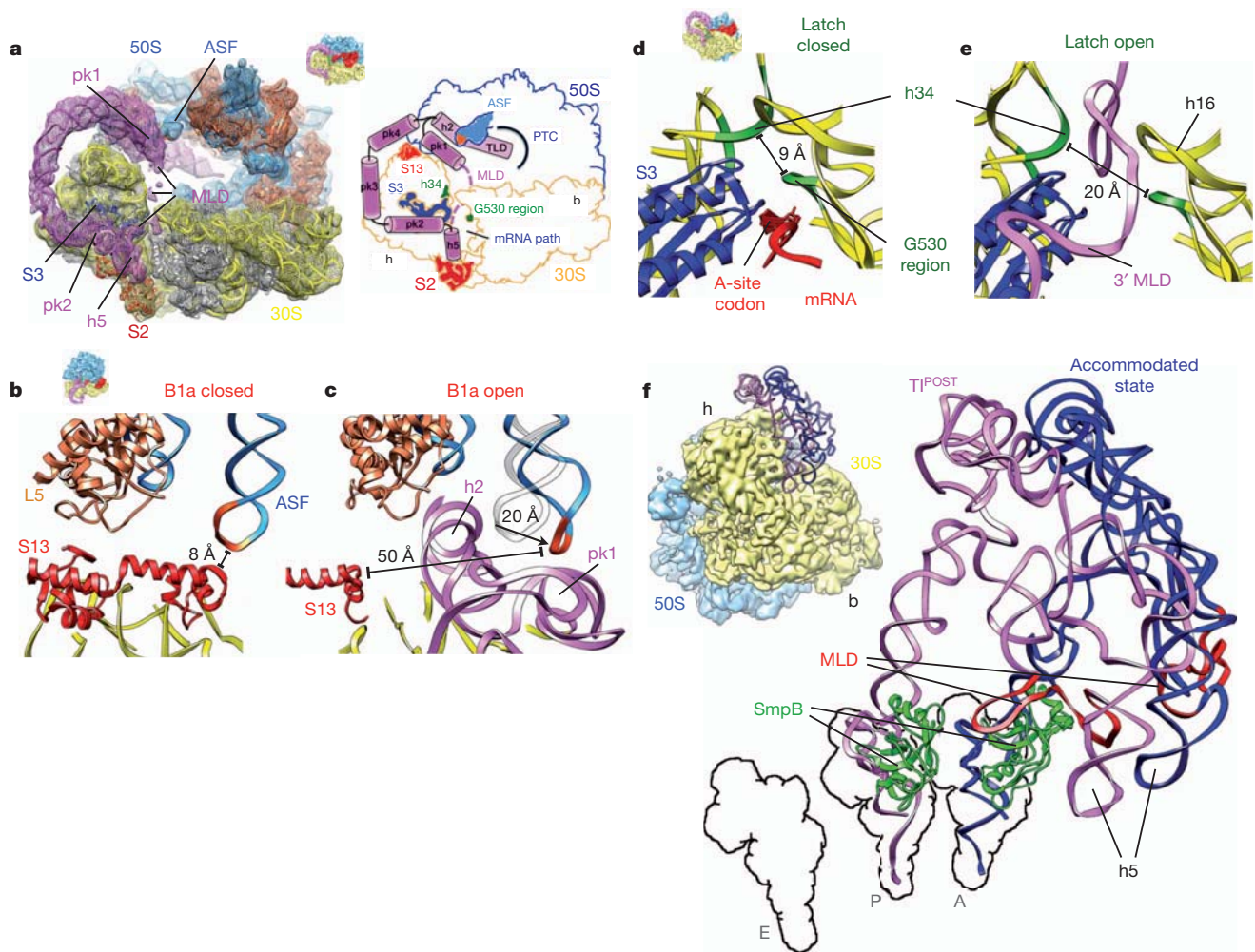


Figure 4 | Structural basis of MLD loading. **a**, Mesh representation (left) of the cryo-EM map with docked models (coloured as in Fig. 2a) and a schematic (right) of the observed $T1^{POST}$ illustrating the ribosome–tmRNA interaction network. **b**, **c**, Cartoon representation of the intersubunit bridge B1a, which is created between the ASF and S13 (red), in the closed form²¹ and the open form after passage of tmRNA-h2. **d**, **e**, Cartoon representation of the 30S latch

(green), which is created between h34 and the G530 region in the closed form²¹ with bound mRNA (red) and the opened form before MLD loading. **f**, Models of the tmRNA–SmpB complex in the accommodated state (blue) and $T1^{POST}$ (pink) shown from the solvent side of the 30S subunit (top left) and indicating the positions of the TLD in the A, P and E sites.

mRNA entry channel is the 30S latch, which is formed by h34 and the G530 (guanosine at position 530) region of 16S rRNA in the 30S head and body, respectively (Fig. 4d). Importantly, the large head movement of the TI^{POST} opens the latch and creates a 20 Å gap between h34 and the G530 region (Fig. 4e). This is reminiscent of eukaryotic translation initiation, in which latch opening is facilitated by the initiation factors EIF1 and EIF1A²² or by internal ribosomal site RNAs^{23,24}. The tmRNA arc wraps around the head and spans the MLD between pk1, located on the intersubunit side, and pk2/tmRNA-h5, located on the solvent side (Figs 1c and 4a). Interestingly, during TI^{POST} formation, both MLD anchor structures — pk1 and tmRNA-h5 — have moved in a direction that pulls the MLD near to the opened latch.

Fragmented densities of the MLD allow the tentative assignment of some parts of its 3' end that interact with S3 and h16 (Fig. 4a, e and Supplementary Table 2). The tmRNA domains that neighbour the 3' MLD (that is, tmRNA-h5 and pk2) still interact with S3 after the translocation reaction⁶. These structural data suggest how MLD loading is accomplished during the repositioning of the 30S head. The backward movement of the head, especially the backward tilting, will cause a displacement of the S3 protein by 20 Å compared with the canonical POST complexes¹⁷. By staying associated with S3, the MLD will be directly guided through the opened latch into the mRNA path.

It seems that MLD positioning by its neighbouring pk1 domain is also crucial for the *trans*-translation process. Deletion of pk1 or its replacement by a structured but shortened RNA element shuts down the tmRNA tagging activity *in vitro* and *in vivo*^{25,26}, and mutations in the ASF correlate with hampered tmRNA function²⁷. Thus, the present structure provides an explanation for the mechanism of MLD loading and suggests that the rotational head movement links TLD translocation to MLD loading. Importantly, assigning a specific functional role to the tmRNA domains located on the solvent side of the 30S subunit (pk2 and tmRNA-h5) provides a rationalization for the extended and complex structure of the tmRNA.

METHODS SUMMARY

Ala-tmRNA-SmpB-EF-Tu-GTP was delivered to ribosomes with a truncated mRNA at the A site and an fMet-tRNA^{fMet} at the P site. The resultant PRE complex was translocated by EF-G in the presence of fusidic acid to stall EF-G on the ribosome. The prepared sample was flash frozen and imaged using a Tecnai G² Polara electron microscope (FEI). Digitized cryo-EM data were processed using multiparticle refinement protocols implemented in the program SPIDER¹⁵. To facilitate the modelling of tertiary interactions within full-length tmRNA, homology models of several pseudoknot-typical motifs, such as triple helices or stem-loop-stem structures, were taken from high-resolution structures of RNA pseudoknots (Protein Data Bank IDs 1RNK²⁸, 1E95 (ref. 29) and 1YMO³⁰) (Supplementary Table 1).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 November 2011; accepted 2 March 2012.

Published online 6 May 2012.

- Keiler, K. C., Waller, P. R. & Sauer, R. T. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**, 990–993 (1996).
- Moore, S. D. & Sauer, R. T. The tmRNA system for translational surveillance and ribosome rescue. *Annu. Rev. Biochem.* **76**, 101–124 (2007).
- Valle, M. *et al.* Visualizing tmRNA entry into a stalled ribosome. *Science* **300**, 127–130 (2003).
- Kaur, S. *et al.* Cryo-EM visualization of transfer messenger RNA with two SmpBs in a stalled ribosome. *Proc. Natl Acad. Sci. USA* **103**, 16484–16489 (2006).
- Weis, F. *et al.* Accommodation of tmRNA-SmpB into stalled ribosomes: a cryo-EM study. *RNA* **16**, 299–306 (2010).
- Weis, F. *et al.* tmRNA-SmpB: a journey to the centre of the bacterial ribosome. *EMBO J.* **29**, 3810–3818 (2010).
- Fu, J. *et al.* Visualizing the transfer-messenger RNA as the ribosome resumes translation. *EMBO J.* **29**, 3819–3825 (2010).

- Gutmann, S. *et al.* Crystal structure of the transfer-RNA domain of transfer-messenger RNA in complex with SmpB. *Nature* **424**, 699–703 (2003).
- Bessho, Y. *et al.* Structural basis for functional mimicry of long-variable-arm tRNA by transfer-messenger RNA. *Proc. Natl Acad. Sci. USA* **104**, 8293–8298 (2007).
- Shpanchenko, O. V. *et al.* Stepping transfer messenger RNA through the ribosome. *J. Biol. Chem.* **280**, 18368–18374 (2005).
- Petry, S. *et al.* Crystal structures of the ribosome in complex with release factors RF1 and RF2 bound to a cognate stop codon. *Cell* **123**, 1255–1266 (2005).
- Schuette, J. C. *et al.* GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J.* **28**, 755–765 (2009).
- Dorner, S., Brunelle, J. L., Sharma, D. & Green, R. The hybrid state of tRNA binding is an authentic translation elongation intermediate. *Nature Struct. Mol. Biol.* **13**, 234–241 (2006).
- Ermolenko, D. N. & Noller, H. F. mRNA translocation occurs during the second step of ribosomal intersubunit rotation. *Nature Struct. Mol. Biol.* **18**, 457–462 (2011).
- Loerke, J., Giesebrecht, J. & Spahn, C. M. Multiparticle cryo-EM of ribosomes. *Methods Enzymol.* **483**, 161–177 (2010).
- Dunkle, J. A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332**, 981–984 (2011).
- Gao, Y. G. *et al.* The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* **326**, 694–699 (2009).
- Jin, H., Kelley, A. C. & Ramakrishnan, V. Crystal structure of the hybrid state of ribosome in complex with the guanosine triphosphatase release factor 3. *Proc. Natl Acad. Sci. USA* **108**, 15798–15803 (2011).
- Ratje, A. H. *et al.* Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* **468**, 713–716 (2010).
- Spiegel, P. C., Ermolenko, D. N. & Noller, H. Elongation factor G stabilizes the hybrid-state conformation of the 70S ribosome. *RNA* **13**, 1473–1482 (2007).
- Yusupova, G. *et al.* Structural basis for messenger RNA movement on the ribosome. *Nature* **444**, 391–394 (2006).
- Passmore, L. A. *et al.* The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol. Cell* **26**, 41–50 (2007).
- Spahn, C. M. *et al.* Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science* **291**, 1959–1962 (2001).
- Spahn, C. M. *et al.* Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell* **118**, 465–475 (2004).
- Nameki, N., Tadaki, T., Himeno, H. & Muto, A. Three of four pseudoknots in tmRNA are interchangeable and are substitutable with single-stranded RNAs. *FEBS Lett.* **470**, 345–349 (2000).
- Wower, I. K., Zwieb, C. & Wower, J. *Escherichia coli* tmRNA lacking pseudoknot 1 tags truncated proteins *in vivo* and *in vitro*. *RNA* **15**, 128–137 (2009).
- Crandall, J. *et al.* rRNA mutations that inhibit transfer-messenger RNA activity on stalled ribosomes. *J. Bacteriol.* **192**, 553–559 (2010).
- Shen, L. X. & Tinoco, I. Jr. The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.* **247**, 963–978 (1995).
- Michiels, P. J. *et al.* Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J. Mol. Biol.* **310**, 1109–1123 (2001).
- Theimer, C. A., Blois, C. A. & Feigon, J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* **17**, 671–682 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Muhs, T. Budkevich, M. Sommer and A. Korostelev for discussions. The strain for tmRNA overexpression and the plasmid for SmpB were provided by A. Muto. We also thank J. Burger for his assistance during the cryo-EM data collection, E. Einfeldt for tRNA preparation and R. Albrecht for the preparation of the reassociated 70S ribosomes. This work was supported by grants from the Deutsche Forschungsgemeinschaft DFG (SP 1130/2-1 to C.M.T.S. and K.H.N., and SFB740 to C.M.T.S.), the German Academic Exchange Service (D/09/42768 to K.R.) and the European Union and Senatsverwaltung für Wissenschaft, Forschung und Kultur Berlin (UltraStructureNetwork, Anwenderzentrum) and the Alexander-von-Humboldt grant (GAN 1127366 STP-2 to H.Y.).

Author Contributions H.Y., D.W., M.P., P.L., Y.T. and O.S. established the *in vitro* system for the reconstitution of ribosome–tmRNA complexes. H.Y. prepared the 70S–tmRNA–SmpB–EF-G complex. D.J.F.R. and T.M. collected the cryo-EM data. D.J.F.R., J.L. and C.M.T.S. carried out the image processing. D.J.F.R., K.R., P.S. and C.M.T.S. carried out the modelling of the full-length tmRNA. D.J.F.R., H.Y., K.R., K.H.N. and C.M.T.S. discussed the results and wrote the paper.

Author Information The electron density map and model of the 70S–tmRNA–SmpB–EF-G complex have been deposited in the 3DEM database with the accession number EMD-5386 and in the Protein Data Bank database with accession numbers 3J18 (30S and ligands) and 3J19 (50S). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.M.T.S. (christian.spahn@charite.de).

METHODS

Buffers. The buffers used were binding buffer A10 (20 mM HEPES-KOH, pH 7.6 (0 °C), 10 mM magnesium acetate, 150 mM ammonium acetate, 4 mM β -mercaptoethanol, 2 mM spermidine and 0.05 mM spermine) and binding buffer B4.5 (20 mM HEPES-KOH, pH 7.6 (0 °C), 4.5 mM magnesium acetate, 150 mM ammonium acetate, 4 mM β -mercaptoethanol, 2 mM spermidine and 0.05 mM spermine).

RNA and protein purification. The sequence of the mRNA was 5'-GGC AAGGAGGUAAAAAUGUA-3', which was purchased from IBA. It contained a strong Shine-Dalgarno sequence (underlined) and the initiation AUG codon (bold). The tmRNA was overexpressed in *E. coli* and purified from a denaturing gel³¹. His-tagged AlaRS, SmpB, EF-Tu and EF-G from *E. coli* were overexpressed in *E. coli* using the T7 expression system and isolated with Ni-NTA agarose (QIAGEN). SmpB was further purified by chromatography on a Superose 12 column (GE Healthcare). The tmRNA was heated at 70 °C for 2 min in binding buffer A10, cooled to room temperature for 20 min and kept on ice until use.

Complex preparation, cryo-EM data collection and processing. Stalled ribosomes were formed at 37 °C by incubation for 10 min in binding buffer B4.5 in a 15 μ l reaction volume containing 10 pmol reassociated 70S, 80 pmol mRNA and 20 pmol [³H]fMet-tRNA. The tmRNA alanylation was performed at 37 °C for 20 min in binding buffer A10, in a 10 μ l reaction containing 10 pmol tmRNA, 300 pmol [¹⁴C]alanine, 20 pmol SmpB, 15 pmol EF-Tu, 50 pmol AlaRS, 4.5 mM ATP, 2.25 mM GTP, 7.5 mM acetylphosphate and 100 μ l ml⁻¹ per assay S-100 fraction. The tmRNA mixture and the stalled ribosome mixture were mixed and incubated at 37 °C for 10 min. Then, 30 pmol EF-G and fusidic acid (to a final concentration of 1 mM) were added, and the mixture was incubated at 37 °C for 10 min. Reaction mixtures were separated in aliquots for both cryo-EM analysis and checking the incorporation of [¹⁴C]alanine-tmRNA into the stalled ribosome complex (70S-[³H]fMet-tRNA-mRNA) by puromycin reaction³². Both amino acids were carried by 45.6% of ribosomes.

Subsequently, the stalled ribosome complex was diluted to a final concentration of 30 nM and flash frozen in liquid ethane on carbon-coated grids (Quantifoil) using a Vitrobot device (FEI). A small initial data set was collected automatically using the Leginon system³³ on a 120 kV Tecnai Spirit transmission electron microscope equipped with a 2k Eagle charge-coupled device camera (FEI). High-resolution micrographs were recorded under low-dose conditions on film (Kodak SO-163) using a Tecnai G² Polara (FEI) microscope operating at 300 kV and a nominal magnification of 39,000. Micrographs were collected at an underfocus that ranged from 2.5 to 5 μ m. The resultant micrographs were digitized on a drum scanner (Heidelberger Druckmaschinen) with a pixel size of 1.26 Å on the object scale. After visual inspection and evaluation of the micrographs, particles were preselected using SIGNATURE³⁴ and boxed out. The complete data set contained 728,424 boxed particle images derived from 302 micrographs. SPIDER³⁵ was used for all further image-processing steps. Multiparticle refinement was carried out as described previously^{15,19,36}. Particle images belonging to subpopulations of particles that did not contain the two large

ligands, tmRNA and EF-G, were selected out and not considered for further analysis.

The final cryo-EM reconstruction of the 70S-tmRNA-EF-G complex from 68,842 particles had a resolution of 8.3 Å using the standard 0.5 FSC criteria (Supplementary Fig. 1). Rigid-body docking of X-ray structures of the ribosomal 50S subunit (Protein Data Bank (PDB) ID 3R8S)¹⁶, 30S head, 30S body (both PDB ID 3R8N)¹⁶, tRNA^{fMet} (PDB ID 3ZVO)¹⁸ EF-G (PDB ID 2WRI)¹⁷, TLD and SmpB (PDB ID 2CZJ)⁹ was performed using UCSF Chimera³⁷. To facilitate modelling of the tertiary interactions within the additional tmRNA regions, homology models of several pseudoknot-typical motifs, such as triple helices or stem-loop-stem structures were taken from high-resolution structures of RNA pseudoknots from the mouse mammary tumour virus (PDB ID 1RNK)²⁸, simian retrovirus 1 (PDB ID 1E95)²⁹ or human telomerase (PDB ID 1YMO)³⁰. Homology models of bulges within RNA helices were taken from ribosomal RNA structures (PDB ID 2J01 and 2AW4)^{38,39} (Supplementary Table 1).

Modelling of tmRNA and energy minimization. The tmRNA model was further refined to improve the geometry of the nucleotides between the pseudoknots and to resolve steric conflicts between individual residues. For the optimization, double-stranded fragments of 6–12 nucleotides were relaxed in simulations enforcing Watson-Crick pairs in defined positions using the MacroMolecule Builder program⁴⁰. The relaxed fragments were inserted and the backbone at the insertion points optimized with ModRNA software⁴¹. The identification of Watson-Crick pairs before and after refinement was carried out with RNAView⁴² and ModRNA.

- Felden, B. *et al.* Probing the structure of the *Escherichia coli* 10Sa RNA (tmRNA). *RNA* **3**, 89–103 (1997).
- Márquez, V. *et al.* Maintaining the ribosomal reading frame: the influence of the E site during translational regulation of release factor 2. *Cell* **118**, 45–55 (2004).
- Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
- Chen, J. Z. & Grigorieff, N. SIGNATURE: a single-particle selection system for molecular electron microscopy. *J. Struct. Biol.* **157**, 168–173 (2007).
- Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
- Penczek, P. A., Frank, J. & Spahn, C. M. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. *J. Struct. Biol.* **154**, 184–194 (2006).
- Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
- Schuwirth, B. S. *et al.* Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).
- Flores, S. C., Wan, Y., Russell, Y. & Altman, R. B. Predicting RNA structure by multiple template homology modeling. *Pacif. Symp. Biocomput.* **2010**, 216–227 (2010).
- Rother, M., Rother, K., Puton, T. & Bujnicki, J. M. ModRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.* **39**, 4007–4022 (2011).
- Yang, H. *et al.* Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.* **31**, 3450–3460 (2003).